A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Clinical and Statistical Prediction: A Reformulation and Some New Data.... 1
Robert R. Holt

Social Influence Phenomena in a Task-Oriented Group Situation..... 13
J. S. Kidd

Personality and Psychopathology..... 18
E. I. Burdock, Samuel Sutton, and Joseph Zubin

The Psychodynamics of Social Mobility in Adolescent Boys..... 31
Elizabeth Douvan and Joseph Adelson

The Relation of Achievement Motivation to Problem-Solving Effectiveness. 45
Elizabeth G. French and Francis H. Thomas

The Effect of Brief Social Deprivation on Behaviors for a Social Reinforcer. 49
Jacob L. Gewirtz and Donald M. Baer

Preference for Delayed Reinforcement: An Experimental Study of a Cultural Observation..... 57
Walter Mischel

Empathy or Behavioral Transparency?..... 62
Uriel G. Foa

Continued on inside front cover

Volume 56 JANUARY 1958 Number 1

CONTENTS—*Continued*

Rigidity, Insecurity, and Stress 67
Leonard H. Ainsworth

Intelligence and Emotional Disturbance: Evidence From Army and Veterans Administration Records 75
John B. Miner and James K. Anderson

Influencing Ethnocentrism in Small Discussion Groups Through a Film Communication 82
Leonard L. Mitnick and Elliot McGinnies

Some Factors in Self-Disclosure 91
Sidney M. Jourard and Paul Lasakow

Two Forms of Social Conformity: Acquiescence and Conventionality 99
Halla Beloff

The Role of Motivation in Psychological Stress 105
William Vogel, Robert W. Baker, and Richard S. Lazarus

Interpersonal Preference and Cognitive Organization 113
Nathan Kogan and Renato Tagiuri

CASE REPORT

Social Behavior of Young Children in the *Kibbutz* 117
Helen Faigin

CRITIQUE AND NOTES

Two Determinants of Conformity: Age of Independence Training and Achievement 130
Allen M. Krebs

The Scale Values of Statements of Attitude Toward the Law After a 25-Year Interval 131
John E. Milholland

Motivational Factors in the Violation of a Prohibition 132
Don L. Kinsbrell and Robert R. Blake

The Achievement Motive and Task Performance: A Replication 133
Robert C. Birney

Two Components of Assumed Similarity 135
Elizabeth Alfert

Neuroticism, Extraversion, and the Perception of the Vertical 139
Ronald Taft and James Coventry

On Authoritarianism and Acquiescence: An Added Note to Bass and Messick and Jackson 141
Fred N. Kerlinger

Volume 56 JANUARY, 1958 Number 1

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER M. BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

GARDNER LINDZEY
University of Minnesota

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Circulation Manager: HELEN ORR
Editorial Assistant: BARBARA B. CUMMINGS

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1958 by the American Psychological Association, Inc

CLINICAL *AND* STATISTICAL PREDICTION: A REFORMULATION AND SOME NEW DATA[1]

ROBERT R. HOLT

Research Center for Mental Health, New York University

THE controversial discussions started a few years ago by Meehl's tightly packed little book, *Clinical vs. Statistical Prediction* (10), still continue—especially among graduate students in psychology, most of whom have to read it. Clinical students in particular complain of a vague feeling that a fast one has been put over on them, that under a great show of objectivity, or at least bipartisanship, Professor Meehl has actually sold the clinical approach up the river. The specific complaints they lodge against the book are, in my opinion, mostly based on misinterpretations, wishful thinking, or other errors, yet I have felt for some time that there was something valid in the irrational reaction without knowing why.

What I propose to show here is that clinicians do have a kind of justified grievance against Meehl, growing out of his formulation of the issues rather than his arguments, which are sound. Finally, I want to offer a slightly different approach to the underlying problems, illustrated by some data. It may not quite make the lion lie down with the lamb, but I hope that it will help us all get on with our business, which is the making of a good science and profession.

THE ISSUES RESTATED

Meehl's book contains a review of the controversy, a logical analysis of the nature of clinical judgment, a survey of empirical studies, and some conclusions. I am not going to go into his treatment of the logical issues and his psychological reconstruction of clinical thinking; for the most part, I agree with this part of the book and consider it a useful contribution to methodology. I want to focus rather on his conception of what the issues are in the controversy, on his treatment of the evidence, and on some of his conclusions.

[1] The research reported in this paper was supported by the Veterans' Administration, the New York Foundation, and the Menninger Foundation, and was carried out jointly by Dr. Lester Luborsky and the author, with the collaboration of Drs. Wm. R. Morrow, David Rapaport, and S. K. Escalona.

Many issues make better reading when formulated as battles, and the field of the assessment and prediction of human behavior has not lacked for controversy-loving gauntlet-flingers. The sane and thoughtful voices of Horst and his collaborators, urging compromise and collaboration (5), have been shouted down by the warcries of such partisans as Sarbin (14) on the actuarial side and Murray (12) on the clinical or (as he put it) organismic. Meehl approached the problem with a full awareness of the feelings on both sides, and apparently with the hope that the therapeutic ploy of bringing them all out into the open at the beginning would enable him to discuss the issues objectively.

In a recent discussion of the stir his book has raised, Meehl has expressed surprise and dismay (11) that his effort to take a balanced and qualified position has led so many people to misunderstand him as claiming that clinical prediction has been proved worthless. Yet he is not blameless; by posing the question of clinical *vs.* statistical prediction, he has encouraged two warring camps to form. This in turn makes it appear all the more compellingly that there *are* two clear-cut types of prediction to be compared.

The root difficulty, I believe, lies in Meehl's acceptance of *clinical* and *actuarial* as concepts that can without further analysis be meaningfully applied to a variety of predictive endeavors of an experimental or practical sort. Accepting them as valid types, he can hardly do anything other than pit one against the other and try to decide what is the proper sphere of exercise for each. But the terms in this antithesis mean many things; they are constellations of parts that are not perfectly correlated and can be separated.

The issue cannot therefore be sharply drawn so long as we speak about anything as complex as "clinical prediction" or "the clinical method." Rather, I think the central issue is the *role of clinical judgment in predicting human behavior*. By clinical judgment here, I mean nothing more complicated than the problem-

solving or decision-reaching behavior of a person who tries to reach conclusions on the basis of facts and theories already available to him by thinking them over.

Let us make a fresh start, therefore, by examining the logical structure of the predictive process with an eye to locating the points where clinical judgment may enter. The following five-step process is idealized, and in practice some of the steps are more or less elided, but that does not hurt this analysis.

First, if we are to predict some kind of behavior, it is presupposed that we acquaint ourselves with what we are trying to predict. This may be called job analysis or the study of the criterion. Perhaps those terms sound a little fancy when their referent is something that seems so obvious to common sense. Nevertheless, it is surprising how often people expend a great deal of time and effort trying to predict a kind of behavior about which they know very little and apparently without even thinking that it might help if they could find out more. Consider the job of predicting outcome of flight training, for example. Many attempts to predict passing or washing out have been made by clinicians without any direct experience in learning to fly, without any study of flight trainees to see what they have to *do* in order to learn how to fly a plane, or of the ways they can fail to make the grade (cf. 4).

There is a hidden trick in predicting something like success in flight training, because that is not itself a form of behavior. It is an outcome, a judgment passed by someone on a great deal of concrete behavior. The same is true for grades in college, success in any type of treatment, and a host of other criteria that are the targets in most predictive studies. Because it is hidden by the label, there is a temptation to forget that the behavior you should be trying to predict exists and must be studied if it is to be rationally forecast. In the highly effective pilot selection work carried out by psychologists during the war, careful job analyses were an important step in the total predictive process and undoubtedly contributed a good deal to the over-all success.

This first stage is hardly a good point at which to try to rely on clinical judgment. The result is most likely to be that guesses, easy and arbitrary assumptions, and speculative extrapolations will attempt to substitute for real information. And no matter how remarkable clinical judgment may sometimes be, it can never create information where there is none.

The *second* logical step is to decide what intervening variables need to be considered if the behavior is to be predicted. As soon as we get away from the simplest kind of prediction—that someone will continue to act the way he has been acting, or that test behavior A will continue to correlate (for an unknown reason) with criterion behavior B—we have to deal with the inner constructs that mediate behavior and the determining situational variables as well. You cannot make a rational choice of the kind of information you will need to have about a person to make predictions without some intervening variables, though they may remain entirely implicit. At this point, judgment enters—always, I think, though it may be assisted by empirical study. The best practice seems to be to give explicit consideration to this step, and to supply judgment with as many relevant facts as possible. This means studying known instances, comparing people who showed the behavior in question with others who in the same situation failed to.

All too often, when the problem of intervening variables is considered at all, it is handled by unaided clinical judgment. For example, in the Michigan project on the selection of clinical psychologists (7), a good many personality variables were rated, but there was no previous work highlighting the ones that might be related to success as a clinical psychologist. It was left up to each judge to form his own conception (from experience, theory, and guess) about what qualities mattered most. Again, this puts a greater burden on clinical judgment than it should reasonably be asked to bear. Yet some clinicians seem to have the mistaken notion that they are being false to their professional ideals if they stir from their armchairs at this point; nothing could be further from the best in clinical tradition, which is unashamedly empirical.

Third, it is necessary to find out what types of data afford measures or indications of the intervening variables, and can thus be used to predict the criterion behavior. If a good job has been done of the preceding step, it may be possible to rely entirely on judgment to make the preliminary selection of appropriate means

of gathering predictive data. For example, if a job analysis and study of persons who have done well at the performance in question both suggest that verbal intelligence and information of a certain type are the main requisites, it would be easy to make good guesses about appropriate instruments to provide the predictive data. I use the word "guesses" deliberately, however, to emphasize the fact that judgment can do no more than supply hypotheses; it cannot substitute for an empirical trial to see whether in fact and under the conditions of this particular study the likely-looking instruments do yield data that predict the criterion.

Notice that almost any actuarial predictive system presupposes carrying through this step. If there is to be an actuarial table, one has to collect great numbers of cases to determine the success frequencies for each cell in the table; if a regression equation is to be used, there must be a preliminary study to fix the beta weights. Unfortunately, it is possible to work clinically *without* first getting an empirical check on one's hypotheses about likely-seeming instruments. At the risk of boring you, I repeat: there simply is no substitute for empirical study of the actual association between a type of predictive data and the criterion. Just as judgment is indispensable in forming hypotheses, it cannot be used to test them.

Perhaps this caution seems misplaced. Do I seem to be urging that you should first *do* a predictive study before embarking on one? I am. That is exactly what happens in actuarial prediction: the formula or table being pitted against judgmental prediction is typically being *cross*-validated, while in none of the studies Meehl cites were the clinical predictions under test being cross-validated. This alone is a major reason to expect superior performance from the actuarial predictions, and again it is a disadvantage under which the clinician by no means has to labor.

The next step, the *fourth* one, is to gather and process the data to give measures of the intervening variables. Meehl clearly recognizes that at this point clinical judgment either may play a large role or may be minimized. At one extreme, the data-yielding instrument may be a machine, a gadget like a complex coordination tester, which automatically counts errors and successes and makes a cumulative record of them. The resulting numbers may be directly usable in a regression equation without the intervention of anyone more skilled than a clerk. At an intermediate level, scoring most psychological tests requires a modicum of clinical judgment, though a high degree of reliability may be attained. At the other extreme is the interview; a great deal of clinical judgment is needed to convert the raw data into indices of the constructs the interviewer wants to assess.

It is easily overlooked that judgment needs the help of empirical study in this phase of the work too. The clinician's training supplies this empirical base in large part, but when he is using a familiar instrument to measure unusual intervening variables, or when he is working with an unfamiliar instrument, judgment grows fallible, and it is no more than prudent to piece it out by careful study of the same kind of predictive data on known subjects on whom the intervening variables have been well assessed independently.

The *fifth* and final step is the crucial one: at last the processed predictive data are *combined* so as to yield definite predictions in each case. The job can be done by clinical judgment, or it can be done by following a fixed rule (an actuarial table or regression equation) in a mechanical way. That much is clear; indeed, this is the locus of Meehl's main interest. I am taking it as granted that a clinician often integrates data in a different way than a statistician—as Meehl says, by performing a creative act, constructing a model of the person from the given facts put together with his theoretical understanding and thus generating perhaps a new type of prediction from a pattern he has never encountered before. We are all curious to know how well good clinicians can do it, and wonder if actuarial combination of data can do as well or better.

But it now seems plain that Meehl has been *too much* interested in this last stage, and as a result has neglected to pay enough attention to the way the earlier aspects of the predictive process were handled in the studies he has reviewed. Here I want to state my main critical point: *If two attempts to predict the same behavior differ significantly in the role played by clinical judgment as against actual study of the facts in one or more of the four earlier parts of the predictive process, a comparison of the suc-*

cesses of the two attempts can tell us nothing definite about the effectiveness of clinical judgment at the final, crucial stage. For this reason, in none of the 20 studies Meehl cites were the comparisons pertinent to the point. Particularly at the vital third step, the predicting statisticians have had the advantage of having previously studied the way their predictive data are related to the criterion; the clinicians have not.

If your reaction is, "So much the worse for the clinicians; nobody stopped them," I am afraid you are thinking about a different question from the one Meehl has raised. If the issue were whether some clinicians have made themselves look foolish by claiming too much, then I should agree: these studies show that they have, and unhappily, they have brought discredit on clinical methods generally. But the studies cited by Meehl and more recently by Cronbach in the *Annual Review of Psychology* (2) unfortunately have too many flaws at other points to tell us what clinical judgment can or cannot do as a way of combining data to make predictions. It is as if two riflemen were having a target match, but one took a wrong turn on the way to the shoot, never showed up, and lost by default. He demonstrated himself to be a poor driver, perhaps, but we never found out how well he could shoot, which is what we really wanted to know.

The other point I want to make in connection with the five-step analysis of the predictive process is this: Since there are so many ways in which clinical judgment can enter, for better or for worse, it makes little sense to classify every attempt to predict behavior on one side or the other of a simple dichotomy, clinical vs. statistical. There can be many types of clinical and actuarial combinations, and many are in fact found in Meehl's mixed bag.

For purposes of exposition, I should like to suggest an only slightly extended typology. Extracting from the best actuarial studies those parts of their procedure during the first four steps that are simply the application of common sense and the scientific method, I propose that we make it quite plain that these can be separated from actuarial prediction at the final step by creating a third type. Thus we should have:

Type I. *Pure actuarial:* Only objective data are used to predict a clear-cut criterion by means of statistical processes. The role of judgment is held to a minimum, and maximal use is made of a sequence of steps exemplified in the most successful Air Force studies in selecting air crew personnel (job analysis, item analysis, cross-validation, etc.).

Type II. *Naive clinical:* The data used are primarily qualitative with no attempt at objectification; their processing is entirely a clinical and intuitive matter, and there is no prior study of the criterion or of the possible relation of the predictive data to it. Clinical judgment is at every step relied on not only as a way of integrating data to produce predictions, but also as an alternative to acquaintance with the facts.

Type III. *Sophisticated clinical:* Qualitative data from such sources as interviews, life histories, and projective techniques are used as well as objective facts and scores, but as much as possible of objectivity, organization, and scientific method are introduced into the planning, the gathering of the data, and their analysis. All the refinements of design that the actuarial tradition has furnished are employed, including job analysis, pilot studies, item analysis, and successive cross-validations. Quantification and statistics are used wherever helpful, but the clinician himself is retained as one of the prime instruments, with an effort to make him as reliable and valid a data-processor as possible; and he makes the final organization of the data to yield a set of predictions tailored to each individual case.

If we now re-examined the studies cited by Meehl and Cronbach, we see that most of them have pitted approximations to Type I actuarial predictive designs against essentially Type II naive clinical approaches. It seems hardly remarkable that Type I has generally given better results than Type II; indeed, the wonder should be that the naive clinical predictions have done as well as they have, in a number of instances approaching the predictive efficiency of actuarial systems.

Other studies cited have come closer to comparing Type II with Type III—naive vs. sophisticated clinical prediction instead of clinical vs. statistical. For example, the prognostic studies by Wittman (16) compared predictions of reaction to shock treatment made in a global way at staff conference with a system she devised. But her system used highly judgmental predictive variables, as Meehl himself points out (ranging from *duration of psychosis* to *anal erotic vs. oral erotic*), and they were combined using a set of weights assigned on judgmental, not statistical grounds.[2] What she showed was that a system-

[2] It is true that the weights were applied in the same way for all cases; in this respect, the system deviates from the ideal Type III.

atic and comprehensive evaluation of the thirty items in her scale (all based on previous empirical work) made better predictions of the outcome of shock treatment than global clinical judgments not so organized and guided. A study of movement in family case work by Blenkner came to a very similar conclusion with somewhat different subject matter (1). When social workers rated an initial interview according to their general impressions, they were unable to predict the outcome of the case, whereas when their judgments were organized and guided by means of an outline calling for appraisals of five factors which had been shown in previous studies to be *meaningfully*, not statistically related to the criterion, then these judgmentally derived predictive variables, combined (like Wittman's) in an a priori formula, predicted the criterion quite well. Yet both studies are tallied as proving actuarial predictions superior to clinical.

Meehl's conclusion from his review of this "evidence" is that clinical prediction is an expensive and inefficient substitute for the actuarial method, and one that keeps clinicians from using their talents more constructively in psychotherapy or exploratory research.

The evidence available tells us hardly anything about the relative efficacy of clinical judgment in making predictions. The weight of numbers should not impress us; as long as the studies don't really bear on the issue, no matter how many are marshalled, they still have no weight. Remember the *Literary Digest* poll: many times more straw votes than Gallup used, but a faulty sampling principle, so that piling up numbers made the conclusion less valid as it got more reliable. Moreover, the studies tallied are so different in method, involving varying amounts of clinical judgment at different points (in the "actuarial" instances as well as the "clinical" ones), that they cannot sensibly be added together.

What is fair to conclude, I think, is that many clinicians are wasting their time when they try to fall back on their clinical judgment in place of knowing what they are talking about. They have been guilty of over-extending themselves, trying to predict things they know nothing about, and learning nothing in the process of taking part in what Cronbach calls "horserace experimental designs," in which clinicians and statisticians merely try to outsmart each other. A multiplication of such studies will not advance clinical psychology.

One kind of comparative study might teach us something even though it would be hard to do properly: simultaneous attempts to predict the same criterion from the same data by clinicians and statisticians *who have gone through the same preliminary steps*. As the statistician studies the original group to determine the critical scores for his multiple cutting point formula (or whatever), the clinician will study the configurations of these scores in individuals of known performance. Then we will see how their respective predictive techniques work.

Does it really make sense, however, for both to use the same data and predict the same criterion? A second possibility would be for two otherwise equally sophisticated methods to predict the same criterion, each using the kind of data most appropriate to it. Or, third, the more clinical and the more statistical methods would not predict the same criterion, but each would undertake to predict the kind of behavior it is best suited to, using the most appropriate kind of data.

Doesn't this third proposal abandon experimental controls necessary for intelligible results? To some extent, yes; but one may have to give up some control to avoid absurdity. As long as clinician and statistician are trying to predict the same criterion, the clinician is likely to be working under a severe, though concealed, handicap. The study will usually have been designed by the statistician; that is his business. He will naturally choose the kind of criterion for which *his* methods seem best adapted; indeed, the nature of his method makes it impossible for him to choose the kind of predictive task that would be most congenial to the clinician, such as drawing a multidimensional diagnostic picture of a total personality, or predicting what a patient will do next in psychotherapy. Thus, the statistician takes advantage of the foolish boast of the clinician, "Anything you can do, I can do better," and plans the contest on his own grounds. The clinician ends up trying to predict grade-point average in the freshman year by a "clinical synthesis" of high school grades and an intelligence test. This is a manifest absurdity; under the circumstances, how could the clinician do other than operate

like a second-rate Hollerith machine? If clinical judgment is really to be tested, it must operate on data that are capable of yielding insights. Moreover, it makes hardly any more sense to expect it to grind out numerical averages of course grades than to expect an actuarial table to interpret dreams.

For reasons of this kind, McArthur (9) recently called for studies of the third type just listed, and maintained that there have as yet been no studies in which the clinician has been given a chance to show what he can do on his own terms. I want therefore to present briefly some results from one such attempt: a study in which clinicians tried to predict criteria of their own choosing, using clinical types of data—interviews and psychological tests (mainly projective techniques).[3] Since some preliminary reports of this work have already been quoted as showing the ineffectiveness of clinical predictions (cf. 2), I have a special desire to set the record straight.

VALIDATING NAIVE AND SOPHISTICATED CLINICAL PREDICTIONS: SOME NEW DATA

The project was an effort to improve the methods by which medical men were selected for specialty training in the Menninger School of Psychiatry. It was begun by Dr. David Rapaport, together with Drs. Karl A. Menninger, Robert P. Knight, and other psychiatrists at the Menninger Foundation at the time the Menninger School of Psychiatry was founded 11 years ago. In the late summer of 1947, Dr. Lester Luborsky and I began work on it, and since then we have jointly carried major responsibility for the project although quite a number of other people have made important contributions.

Our work consisted of two predictive studies. Following the terminology suggested above, one used a naive clinical method, while the other was an attempt at a more sophisticated clinical method. The naive clinical design was simple: Psychiatrists and psychologists used their favorite means of assessing personality to forecast the level of future performance of psychiatric residents at the time when they applied to the School. The applicant came to Topeka after some preliminary correspondence, having survived a rough screening of credentials and application forms. He was seen by three psychiatrists, each of whom interviewed him for about an hour, rated his probable success as a resident on a 10-point scale, and made a recommendation: Take, Reject, or Doubtful. The psychologist made similar ratings and recommendations after analyzing results of a Wechsler-Bellevue, Rorschach, and Word Association test. In addition, both psychologists and psychiatrists submitted brief qualitative reports of their appraisals of the man's positive and negative potentialities. All of the data and predictions were turned over to the Admissions Committee, which made the final decision to accept or reject each man.

During the years of the project, from 1946 through 1952, six successive classes of residents were chosen. The first 456 applicants who went through this procedure formed our experimental population (excluding small numbers of Negroes, women, and persons from Latin-American and non-European cultures, since these minorities offered special problems of assessment). A little over 62 per cent of these applicants were accepted by the Committee, but only 238 actually entered the School; 46 changed their minds and went elsewhere. Nevertheless, we kept in touch with them and the 172 rejectees by a mail follow-up questionnaire for several years, so that we have data on certain aspects of their subsequent careers.

The clinicians making the predictions had in some cases had considerable experience in training psychiatric residents, but there was no explicit job analysis or preliminary study of criterion groups. They simply fell to and made their predictions and their decisions.

To test the validity of these clinical decisions, let us use as a criterion, first, whether or not a man passed the certifying examination of the American Board of Psychiatry and Neurology—the criterion set up by the specialty itself. We have this information on all subjects from the lists published by the Board. The Admissions Committee's decisions had a good deal of validity as predictors of this criterion: 71 per cent of the men they voted to accept had passed the Board examination in psychiatry by the end of 1956, while only 36 per cent of the rejected candidates had done so. This difference is significant at better

[3] This study is presented at length in a forthcoming book (3).

than the .01 point. The recommendations made by interviewers (taking them all as a group) and the recommendations of the psychological testers to accept or reject likewise were highly valid predictors of this criterion, significances in both cases also being beyond .001.

It is interesting that the Committee decisions were slightly *better* at predicting both staying in psychiatry and passing the certification examination of the American Board of Psychiatry and Neurology—better than either the psychiatric interviewers or the psychological testers. It is possible, however, that much or all of this apparent superiority is due to the fact that rejection by the committee did discourage a few applicants from seeking training elsewhere.

Data of this kind are encouraging to the people trying to run a school of psychiatry but hard to interpret in a larger context. Who knows but that an actuarial table based on objectively ascertainable facts like grades in medical school, marital status, age, etc. might not have done just as well? We never took such a possibility very seriously, but we did try out a few such objective predictors in our spare time, just out of curiosity. None of them showed any particular promise as a predictor of any criterion taken alone, though it is possible that patterns of them such as an actuarial table uses might have operated a little better than chance.

The criteria on which we spent most time and labor were measures of competence in psychiatric work during the last two years of the three-year residency. Whenever a resident completed a period of time on a particular service, we would interview the staff men, consultants, and others who had directly supervised his clinical work, and get them to rate it quantitatively. The resulting criterion measure had a coefficient of internal consistency above **.9**[4] (for the last few classes), and we have every reason to think it has a great deal of intrinsic validity. We also got the residents to rate each other's work. The reliability of their pooled ratings of over-all competence is also **.9**, and this criterion (which we call "Peers' Evaluations") correlates from **.66** to **.78** with Supervisors' Evaluations.

These criterion judgments enable us to test the validity of the predictive *ratings*. The validity correlations are not exciting, though for the entire group of residents in the Menninger School of Psychiatry they are all significantly better than zero at the one per cent point. Taking the *mean* of the ratings given by the psychiatrists who interviewed an applicant, this predictor correlates *.24* with Supervisors' Evaluations. The predictions of the psychological testers were not significantly better: the validity coefficient is **.27.** There was some fluctuation from class to class, the interviewers' validities varying from exactly .00 to *.52*, the tester's validities from .12 to **.47.** Likewise, the validities of ratings made by individual clinicians vary over the same range: psychiatric interviewers from .01 to **.27** ($N = 93$) or *.47* (significant at only the 5 per cent point because the N was only 13), and psychologists from .20 to **.41** ($N = 40$). At the same time, those individual clinicians all did much better in making the basic discrimination: recommending acceptance of men who actually became psychiatrists.

These correlations are nothing to get excited about, and nothing to be ashamed of either, particularly in view of the restriction of range in this selected, accepted sample. They show that the naive clinical method depends a good deal on the ability of the particular clinician doing the predicting, and that—at least in this study—a pooling of judgment helped make up for the deficiencies of individuals.

Let us turn now to the second experimental design, which I have called a *sophisticated clinical* type of prediction. I shall have to skip lightly over many complicated details, and make things look a little more orderly than they actually were. The design included a job analysis of the work done by psychiatric residents, which was broken down into a few major functions (such as diagnosis, psychotherapy, administration of wards) and 14 more specific aspects of work. Then we attempted to specify attributes of personality that would facilitate or hinder a man in carrying out such work, first by collecting the opinions of persons who had had long experience in training psychiatrists, psychotherapists, or psycho-

[4] Correlations in bold type are significant at the one per cent point, those in italics, at the five per cent point. One-tailed tests are used throughout to test the null hypothesis that the predictor does not correlate *positively* with the criterion.

analysts. The second way we went about it was to make an intensive study of a dozen excellent residents and a dozen who were rated at the bottom by their supervisors. We went over all the original assessment data on them, interviewed them and tested them extensively, trying out many novel approaches and then seeing what discriminated these known extreme groups. Thus, we learned what personological constructs differentiated good from poor residents, and what tests and test indicators gave evidence of these constructs. Hoping to guide clinical judgment in the use of interviews and projective tests, we used the data from these small samples of extremes to help us write detailed manuals on the use of the interview, TAT, Rorschach, and other techniques in the selection of psychiatric residents. The manuals listed discriminating cues, both positive and negative, which were to be summed algebraically. We then made preliminary cross-validations of these manuals (as many as we could) with encouraging results (cf. 8) and revised them after studying predictive successes and failures.

As a last step, we set up another predictive study to submit our manuals to a final cross-validation on a group of 64 subjects and to accomplish several other purposes at the same time. Four psychologists served as judges; each of them scored tests or interviews according to our manuals and also made free clinical judgments based on increasing amounts of data. Two of the judges made such predictive ratings after going through an entire file of assessment data: credentials, intellectual and projective tests, and a recorded and transcribed interview.

How did we make out? Considering first the manuals, only indifferently well. Of the six, two proved worthless (TAT Content and a special projective test); the other four all showed more or less promise, but there was none that yielded consistently significant validities regardless of who used it. Reliability, in terms of scorer agreement, was on the whole not very good, for a good deal of clinical judgment was still demanded. Consider, for example, one TAT cue that worked well for one judge (validity of *.26* against Supervisor's Evaluations of Overall Competence). This cue called for judgments of the *originality* of each TAT story, obviously a matter on which psychologists might fairly easily disagree: scores by Judges I and IV correlated −.04. The validities attained by the manuals for the Interview, Rorschach, Formal Aspects of the TAT, and Self-Interpretation of the TAT were on about the same general level as those from our first, naive clinical design—mostly in the .20's.

Now for the free clinical predictive ratings. When the judge had only one test or an interview to go on he usually added little by going beyond the manual and drawing on his general experience and intuition. Some judges did slightly better with free ratings than with manual scores, some a little worse. At this point, you may wonder at all this exposition for so small a result: barely significant validity coefficients, about the same size as those from a naive clinical approach, despite the attempt to create a sophisticated clinical predictive system that involved many actuarial elements. I believe that the lesson of our findings up to this point is simple: *With an inadequate sample of information about a person, no matter how sophisticated the technique of prediction, there is a low ceiling on the predictive validity that can be attained.* In our experience, even a battery of two or three tests (if exclusively projective), or an interview and a couple of projective techniques, does not give an adequate informational sample to enable clinicians to make very accurate predictions of complex behavioral outcomes over a period of three years.

Look next at the results when experimental judges made their predictions from as complete a body of information about a man as could be assembled at the time he applied. The hard-headed statistical expectation *should* be that validities would at best remain at the same level, and more likely would decline. The widely read preliminary report of the Michigan project on the selection of clinical psychologists (6) reported declining validities as increasing amounts of information were made available to judges (in a design which, for all its complexity, was essentially a naive clinical one). Not so many people have read the full final report (7), where this issue is not discussed; it is necessary to pore over many tables and tally numbers of significant correlations at various stages to find out for oneself that with the final criteria there was a slight *rising* trend in validities of clinical predictions as the

TABLE 1

SOME VALIDITIES FROM SYSTEMATIC CLINICAL ASSESSMENT OF APPLICANTS FOR RESIDENCIES IN THE MENNINGER SCHOOL OF PSYCHIATRY

Predictors:	Validities of Predictors Against Criterion Evaluations of:									
	Overall Competence		Competence in Psychotherapy		Competence in Diagnosis		Competence in Management		Competence in Administration	
	Sup.[a]	Peer[a]	Sup.	Peer	Sup.	Peer	Sup.	Peer	Sup.	Peer
Predictive Ratings										
Judge I: PRT[b]	*.26*	.23	.12	*.26*	.13	.21	*.31*	.00	.20	.10
Judge II: TAT	−.10	−.02	−.16	−.01	−.05	.20	−.08	.01	.04	.11
Judge I: All data	**.57**	**.52**	**.48**	**.55**	**.58**	**.42**	**.52**	*.36*	**.55**	**.42**
Judge II: All data	*.22*	**.48**	.15	**.36**	*.24*	**.42**	.13	*.24*	*.24*	*.27*
Liking Ratings										
Judge I: PRT	*.29*	*.34*	.16	**.35**	*.24*	**.36**	.15	.17	.19	.16
Judge II: TAT	−.02	.13	−.08	.15	.00	*.30*	−.17	−.02	−.14	.10
Judge I: All data	**.58**	**.64**	**.45**	**.58**	**.51**	**.52**	**.52**	**.52**	**.50**	**.46**
Judge II: All data	*.25*	**.49**	.20	**.47**	*.21*	**.56**	.10	*.30*	.18	*.30*

[a] Sup. = Supervisors' Evaluations; Peer = Peers' (Sociometric) Evaluations.
[b] PRT = Picture Reaction Test, a specially devised projective test similar to TAT.
Numbers of cases: For Judge I—Supervisors' Evaluations, PRT: 63, all data: 37
Peers' Evaluations, PRT: 45, all data: 30
For Judge II—Supervisors' Evaluations, TAT: 63, all data: 64
Peers' Evaluations, TAT: 45, all data: 46

amount of information available to the predicting judge was increased.

The same thing is true of our results, but in a dramatic and unmistakable way (see Table 1). Considering only the two judges who went through the entire mass of material, their final free clinical ratings of Over-all Competence correlated **.57** and *.22* with Supervisors' Evaluations and **.52** and **.48** with the sociometric Peers' Evaluations. In considering these correlations, remember that they are attentuated by a significant restriction of range, since all subjects had successfully passed through an Admissions Committee screening which had considerable validity. They have not been corrected for less than perfect reliabilities, either.

An incidental finding is even more remarkable. The predictive analysis was made approximately a year after the assessment data had been gathered. The judges went right through the entire series of cases, making their ratings in a *blind* analysis; names and other identifying data were concealed, and there was no direct contact with the subjects. Nevertheless, judges formed rather vivid impressions from the material, including a feeling of how well they would like each candidate personally. For control purposes, they were required to rate this feeling of *liking*. When we undertook to correlate the liking rating with predictors and criteria so as to partial out this possible source of error, we found that it was the best predictor we had! These ratings of liking by Judges I and II correlated highly with their predictive ratings, but even more highly with Supervisors' Evaluations (**.58** and *.25*) and especially with Peers' Evaluations (**.64** and **.49**).[5] A study of these liking ratings suggests

[5] One consequence of the delay between the gathering of the data and their analysis was that some of the *S*s who entered the Menninger School of Psychiatry became known to the predictive judges, raising the possibility of contamination of their predictions by criterion-relevant knowledge. Despite the fact that the analysis of the assessment material was done "blind," identifying data having been removed or concealed, Judge I fairly often recognized the identity of *S*s at the final stage of analysis. He therefore did not make predictive ratings, which is why his *N*s are so low for this stage. There were a few borderline instances of partial or questionable recognition, however, in which Judge I (four cases) or II (two cases) had some information or misinformation about the subject. If these cases are eliminated, Judge II's validities go up more often than they go down, the range being from a decrease of .16 to an increase of .15. Judge I's validities against Supervisors' Evaluations are negligibly affected, the range of effects extending from a loss of .08 to a gain of .20. Most of his validities against Peers' Evaluations were more seriously affected, however, especially at the final (all-data) stage of analysis, where losses of up to 28 correlation points occurred. On the whole, however, even a very conservative handling of the problem of possibly contaminated cases does not change the essential import of the results: It was still possible for Judge I to obtain four validities of **.50** or higher, and for Judge II to obtain two validities

to us as the most plausible explanation that they differed from our intentional clinical predictions in being somewhat more irrational, affective—perhaps intuitive—reactions to the same data.

In all of the correlations I have been citing, you will perhaps have noticed that one judge consistently did slightly better than the other. This is certainly to be expected. When clinical judgment is the main technique of processing data, there are bound to be differences due to the skill of clinicians in doing this particular job.

Are we justified in citing these few high validities as evidence of what the sophisticated clinical method can do in a study where it is given a chance to prove itself on grounds of its own choosing? I believe that we are. The psychologists who were our Judges I and II were considered to be good but not extraordinary clinicians, certainly no better than the best of the psychologists and psychiatrists who made the "naive clinical" predictions. They differed principally in that they had an adequate sample of data and had been through all the preliminary stages of studying the criterion in relation to the predictors in earlier groups of subjects. Moreover, they used systematic methods of analyzing the data, attempting to record all inferences and working with a set of intervening variables—personality constructs which were directly inferred from the test and interview data, and from which in turn the behavioral predictions were made. In a true sense, their clinical ratings were not naively based on unguided and uncontrolled judgment; they constituted a cross-validation of a whole structure of hypotheses about the determinants of psychiatric performance based on intensive prior study. Even so, our study left a great deal to be desired as a model of the sophisticated clinical approach—particularly on the scores of (*a*) a better job analysis, (*b*) a more broadly based, configurational approach to the design of manuals, and (*c*) a better stabilized criterion (see 3).

By way of contrast, a few more data before a final summing up. You remember that the battery of tests used in the first naive clinical predictive design consisted of the Wechsler-Bellevue, Rorschach, and Word Association tests. (The Szondi was also given, but was usually ignored; and the Strong Vocational Interest Blank was also routinely given, but was never scored in time to be used in the actual assessment for the Admissions Committee). The Rorschachs we gave were of course scored in the conventional way as well as by our special manual, and we thought it might be fun to see how some of the usual Rorschach scores would be related to the criterion. So we tried 14 scores and simple indices (like A%) with one class, and were surprised to find some rather high correlations. We decided therefore to see how a straight statistical-actuarial method of using the Rorschach would perform. Scrutinizing the table of intercorrelations between the Rorschach scores, we chose five of them that promised the greatest chances of success: DR%, number of good M, new F+%, F% and Stereotype% (scored after Rapaport, 13)—the last two because they looked as if they might be good "suppressor variables." A multiple regression equation was worked out to give the best linear combination of these variables to predict Over-all Competence; R for this Class was **.43** ($N = 64$). We noticed, however, that in the regression equation only the first two scores seemed to be playing any appreciable part, and in fact the multiple correlation using only per cent of rare details and number of good M was also **.43**. The other three were dropped from the formula, which was then tested on the Rorschach scores and criterion ratings of the first three classes. As expected, the correlation dropped out of sight on being cross-validated; with the new group of 116 subjects, it was .04.

The Strong Vocational Interest Blank, which (with an intelligence test) gave the best validities in the Kelly-Fiske study, likewise failed to yield any good predictor of competence in psychiatry. Even the special key ("Psychiatrist A") produced by Strong from a statistical analysis of thousands of blanks filled out by diplomates in psychiatry (15) failed to predict any of our criteria at a statistically significant level: no r's as high as .2. This last finding deserves emphasis, because Strong's key was the product of a highly developed statistical technology, had an adequate numerical base, and had every opportunity to show what a pure actuarial method could achieve.

of **.36** or higher. (For a fuller discussion of the problem of contamination in these results, see [3].)

We made a further attempt to combine the best-predicting scores from the tests used in the standard battery into a regression equation. The *R* between Verbal IQ (Wechsler-Bellevue), Lawyer key (Strong VIB), DR% and No. Good M (Rorschach), and Overall Competence was .56 on the original group of 64; cross-validated on 100 cases, it dropped to .13.

SOME PRACTICAL IMPLICATIONS

If we had concentrated on an actuarial rather than a clinical approach and had come up with a simple, objective procedure that had a high and stable level of validity in predicting psychiatric performance, it could have been misused. It might have tempted many psychiatric training centers to adopt a single mold from which would have been cast a generation of psychiatrists, who would have had to meet the problems of the future with a standard set of resources derived from the past. The more successful we are in finding objective, impersonal, and statistical methods of selecting members of a profession in the image of its past leaders, the more rigid will be the pattern into which it is frozen.

For a concrete example, consider Strong's Psychiatrist A key again for a moment. It expresses the pattern of interests held in common by men who were diplomates in psychiatry at the end of the war, most of whom must have trained fifteen to twenty years ago. It should hardly be surprising that residents whose interests most closely approached this pattern tended to have skills as administrators and diagnosticians rather than as psychotherapists. If they had happened to achieve a high correlation with our over-all criterion, it might have helped populate American psychiatry of the 1960's and 1970's with near-replicas of the old state hospital superintendent.

It might be argued, however, that a similar result could have been expected if we had succeeded in providing explicit methods of clinically analyzing other types of data to select psychiatric residents. They too would have been based on a study of men who were successful at one time in history, and would have suffered the same danger of getting out of date. The answer is that even sophisticated clinical prediction never gets quite that rigid. Changes creep in; the result may be that validities gradually regress, or the drift may be determined by valid appraisals of newly important variables. Clinical methods are more flexible than their actuarial counterparts; they *can* be more readily modified by new studies based on observations of developing trends in the criterion. Moreover, valid clinical impressions can be obtained from an intensive study of a few known cases, while it takes large samples to set up or revise an actuarial system. There can be no guarantee that clinical methods *will* be kept up to date, of course, nor that the attempt to do so will not spoil their validities. Any predictive system needs constant overhaul and revalidation.

By sticking with the only capriciously accurate, sporadically reliable, and eminently flexible method of clinical judgment in selecting trainees, psychiatry will at least be able to keep in touch with developments in a growing and changing profession. Moreover, it will be able to maintain a healthy diversity within its ranks. There are many jobs to be done in psychiatry, requiring quite different kinds of men. There must be thoughtful men who like to sit in deep chairs and analyze patients all day long. There must be activists to organize new institutions and give inspirational leadership to groups of colleagues. Psychiatry needs many more men than it has whose main interest is in research and teaching, others to work with broad preventive programs in public health, group therapists, specialists in somatic treatments, and many more varieties of the general species. If the pure actuarial approach were to be seriously applied to psychiatry, it would be necessary to develop a formula for each of many different types of practice and to revise it constantly as new developments created needs for new types of practitioners. To do so would be impossibly expensive and laborious. Psychiatry is well-off, therefore, sticking with a basically clinical approach to assessment and prediction in selecting its members, but trying constantly to make it more scientific.

The important issue, however, is not what method of selecting its members is best for any particular profession, but the relative inertia of actuarial predictive systems and the maneuverability introduced when the generating of predictions is done by clinical judgment. This freedom is a source of weakness as well as

strength; it enables the clinician to fall into errors of many kinds (to which statistical predictions are less subject) and also to adapt himself sensitively to all kinds of changing circumstances. When clinical methods are given a chance—when skilled clinicians use methods with which they are familiar, predicting a performance about which they know something—and especially when the clinician has a rich body of data and has made the fullest use of the systematic procedures developed by actuarial workers, including a prior study of the bearing of the predictive data on the criterion performance, then sophisticated clinical prediction can achieve quite respectable successes. I hope that clinicians will take some heart from our results, but I urge them to refine their procedures by learning as much as possible about statistical prediction and adapting it to their own ends.

To summarize: Meehl failed in his aim to mediate the statistical–clinical quarrel because he defined the issues in a way that perpetuates competition and controversy. The real issue is not to find the proper sphere of activity for clinical predictive methods and for statistical ones, conceived in ideal-type terms as antithetical. Rather, we should try to find the optimal combination of actuarially controlled methods and sensitive clinical judgment for any particular predictive enterprise. To do so, we must strike the right balance between freedom and constraint, a balance which may shift a good deal in one direction or the other depending on the nature of the behavior being predicted. But we can find such balances only if clinically and statistically oriented workers give up contentious, competitive attitudes and seek to learn from each other.

REFERENCES

1. BLENKNER, M. Predictive factors in the initial interview in family casework. *Soc. Serv. Rev.*, 1954, **28**, 65–73.
2. CRONBACH, L. J. Assessment of individual differences. In P. Farnsworth & Q. McNemar (Eds.), *Annual review of psychology*, Vol. VII. Stanford, Calif.: Annual Reviews, 1956. Pp. 173–196.
3. HOLT, R. R., & LUBORSKY, L. *Selecting psychiatric residents*. New York: Basic Books, in press.
4. HOLTZMAN, W. H., & SELLS, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.
5. HORST, P. *et al. The prediction of personal adjustment*. New York: Soc. Sci. Res. Council Bull. 48, 1941.
6. KELLY, E. L., & FISKE, D. W. The prediction of success in the VA training program in clinical psychology. *Amer. Psychologist*, 1950, **5**, 395–406.
7. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor, Mich.: Univer. of Michigan Press, 1951.
8. LUBORSKY, L. L., HOLT, R. R., & MORROW, W. R. Interim report of the research project on the selection of medical men for psychiatric training. *Bull. Menninger Clinic*, 1950, **14**, 92–101.
9. MCARTHUR, C. Clinical versus actuarial prediction. In *Proceedings, 1955 invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1956. Pp. 99–106.
10. MEEHL, P. E. *Clinical vs. statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
11. MEEHL, P. E. Clinical versus actuarial prediction. In *Proceedings, 1955 invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1956. Pp. 136–141.
12. OSS Assessment Staff. *Assessment of men*. New York: Rinehart, 1948.
13. RAPAPORT, D., GILL, M. M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. II. Chicago: Yearbook Publishers, 1946.
14. SARBIN, T. R. A contribution to the study of actuarial and statistical methods of prediction. *Amer. J. Sociol.*, 1943, **48**, 593–603.
15. STRONG, E. K., JR., & TUCKER, A. J. The use of vocational interest scores in planning a medical career. *Psychol. Monogr.*, 1952, **66** (9), No. 341.
16. WITTMAN, M. P. A scale for measuring prognosis in schizophrenic patients. *Elgin Papers*, 1941, **4**, 20–33.

Received April 15, 1957.

SOCIAL INFLUENCE PHENOMENA IN A TASK-ORIENTED GROUP SITUATION[1]

J. S. KIDD[2]

Northwestern University

THE roughly similar results for social influence on psychophysical judgments (Sherif [9]) and on attitudes and values (Crutchfield [3]) seem to justify the use of the simpler and more easily quantified psychophysical judgments to bring important social phenomena into the laboratory for controlled study. The present experiment was designed to explore the effects of several situational variables on the phenomena of social influence in the behavioral context of a task-oriented group. It was assumed for the purposes of this study that the mechanisms involved in social influence would be manifest subsequent to the exposure of the subject (*S*) to group participation and that the behavior to be influenced need not be relevant to the group task. Social influence is defined here as a change in a response following the presentation to *S* of a response standard of apparent social derivation. A change leading to a higher degree of conformity between the response and the standard indicates positive social influence; a change in response in the direction of less conformity is taken as negative social influence.

The three variables tested were the size of the group, the duration of group participation, and the identifiability of the source of the response standard. Group size was selected as a variable in part because of the apparent incompatibility between the findings of Asch (1) and those of Goldberg (4). Asch found an increase in social influence as the size of the (reporting) group increased. Goldberg, using a different experimental approach, was not able to demonstrate this effect. Interest in duration of participation stemmed from empirical work on voluntary social groups by Bovard (2), who found that conformity (social influence) behavior could not be predicted from length of membership. The present study sought to extend Bovard's findings by considering the same variable against the background parameters provided by nonvoluntary groups. Identifiability of the source of the response standard was suggested as an important variable by an interpretation of some of the general conclusions of Hovland, Janis, and Kelly (5) concerning the effect of personal presence in persuasion.

The three experimental hypotheses tested were that positive social influence increases (*a*) as group size increases, (*b*) as duration of participation increases, and (*c*) if the source of the response standard is identifiable rather than remaining relatively anonymous. These hypotheses, while not entirely consistent with the empirical findings previously cited, provide an internally unified approach compatible with much contemporary theory.

As a secondary consideration in this study, task performance measures were obtained from the various sized groups in order to test the hypothesis that task effectiveness increases with group size.

METHOD

Design and Subjects

The variables of group size and duration of participation were each evaluated at three levels. Groups of two, four, and six members were compared. The durations of group participation selected were 1, 2, and 3 hrs. The variables were incorporated in a 3 x 3 factorial design. In order to test the hypothesis regarding the identifiability of the source of the response standard, an additional (nonfactorial) condition was introduced that involved the observation of solitary *S*s.

In the conditions of two- and six-member groups there were 18 *S*s in each cell; in those for four-member groups, 20 *S*s. Fourteen *S*s were employed in nonfactorial condition. In all cases, individuals rather than groups were considered as the responding entities. The total sample *N* of 182 was composed of 100 female and 82 male undergraduate students from beginning social science classes at Northwestern University. Sampling bias was minimized by randomizing the sequence of conditions and by assigning *S*s to conditions in the order in which they volunteered for participation in the experiment.

[1] This study is based upon a dissertation submitted to the graduate school of Northwestern University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is grateful to Drs. Donald J. Lewis and Donald T. Campbell for their patient assistance and sympathetic sponsorship.

[2] Now with the Laboratory of Aviation Psychology, Ohio State University.

TABLE 1
NUMBER OF CASES AND MEANS FOR MEASUREMENT OF SOCIAL INFLUENCE

Group Size	Duration of Contact			Mean[a]
	1 Hr.	2 Hrs.	3 Hrs.	
One member		14		4.00
Two members	18	18	18	5.18
Four members	20	20	20	5.27
Six members	18	18	18	5.54
Mean	5.02	5.63	5.34	

[a] The scale used was from 1.00 through 9.00 with the higher score indicating greater positive social influence.

TABLE 2
ANALYSIS OF VARIANCE OF SOCIAL INFLUENCE SCORES

Source of Variance	*df*	Mean Square	*F*
Group size	2	2.00	.51
Duration of contact	2	5.03	1.28
Interaction	4	6.50	1.65
Within groups error term	159	3.94	

Procedure

All *S*s were first informed that the purpose of the experiment was the determination of work team effectiveness in accomplishing simple tasks. They were told that on completion of the experiment they would be asked to participate in an additional study, but the nature of the second study was not revealed.

The experimental sessions that followed were of 1-hr. duration. Those groups meeting for two or three sessions came together at intervals of 48 hrs. During each session the groups (or individuals) worked on three tasks: an anagram task, a jumbled sentence task, and a jigsaw puzzle task. In the group situation, the task was structured so as to force cooperative activity, by distributing task materials equally through the group and requiring that the group product contain contributions from each group member. In the condition employing single *S*s, all task materials were presented simultaneously. Knowledge of task results was held at an artificial constant value by *E* in order to minimize variability for any "group success" effect (6).

Following the final task session, *S*s were informed that the main experiment was complete. They were asked then to help out in a brief study of a simple psychophysical nature. The *S*s were seated along the back of the experimental room equidistant from the stimulus source and separated from one another by a minimum of seven feet. The room was partially darkened. The *S*s were provided with response slips and were asked to observe a flickering light for five seconds. Their required response was to estimate the number of flickers in that period. When their estimates were complete, *E* collected the response slips and after a short interval returned to each *S* a fresh response slip upon which was written a purported "group average estimate." The *E* called attention to this figure, indicated that it was the average estimate of the group and that it was presented to give "some idea of what other members of your group had judged the correct total to be." The *S*s were then required to observe the flicker again for five seconds and note a second estimate.

In each case, the purported group average estimate was fictitious. A standard factor of .5 or 1.5 (depending on whether the *S*'s original estimate had been higher or lower than the true value) was applied and the result was reported to each *S* individually as the group figure. This procedure was followed in order that the degree of deviation from the response standard would be proportionately the same for each *S*. The situation was structured so that the discrepancies between each *S*'s "group average estimate" score could not be noted.

For *S*s who worked through two sessions of the task phase alone, the instructions at the terminal measurement of social influence were intended to provide a nonidentifiable source for the response standard. Following *S*'s first flicker estimate, he was given a second slip upon which a figure labeled "average estimate" appeared, and informed that this estimate represented his own original estimate averaged in with the estimates of "some other people in this situation." Neither the number nor any other characteristic of the fictitious others was indicated, thus providing a relatively anonymous or nonidentifiable response source.

For all *S*s, a brief description of the true nature of the study was given following the completion of the social influence measurements.

RESULTS

Social Influence

The effects of the two main variables of group size and duration of participation upon the social influence response were evaluated by both analysis of variance and by the nonparametric χ^2 contingency test. The results of both methods of analysis fail to reveal reliable differences attributable to the experimental conditions. A summary of the data and the statistical analyses is presented in Tables 1, 2, and 3.

The original raw data were collected in the form of percentage change in response with algebraic signs to indicate direction of change. The complete raw score distribution was skewed sharply toward the positive side. For purposes of the analysis of variance, a log transformation was applied and the resultant transformed scores were grouped in nine class intervals. The resulting scale of social influence was found to be suitable for analysis.

Tests of the main effects and interactions using analysis of variance were negative. The small trend in favor of greater social influence with increasing group size was not significant.

The results of the analysis of variance were confirmed by the use of χ^2 contingency tests on the raw data, which fell into a natural trichot-

omy: positive social influence, no change, and negative social influence. A 3 x 3 contingency table arrangement was used to test the effects of the two continuous variables. The variable of group size yielded a χ^2 of 4.75 which is not significant. The duration of group participation variable gave a χ^2 of 1.95 which is also not significant.

In order to test the hypothesis regarding identifiability of the source of the response standard, a comparison was made between the influence scores of *S*s who participated as individuals and *S*s in the two-member group condition having the same duration of participation (in this case two task sessions). The difference between means was evaluated by *t* test. The *t* ratio obtained was 4.01 which for 30 *df* is significant beyond the .01 level. Further confirmation was accomplished by pooling the response data for the three conditions of group size (two, four, and six members), and comparing the resultant mean score with the mean score from the condition of low identifiability of the source of the response standard. The *t* ratio obtained in this comparison was 5.14 with 68 *df*, which was again significant beyond the .01 level of probability. In both instances, the difference between the means was in the direction of greater positive social influence when the source of the response standard was identifiable. A summary of *t*-test comparisons is presented in Table 4.

TABLE 3
FREQUENCY ANALYSIS OF SOCIAL INFLUENCE RESULTS

Direction of Social Influence	Group Size			Totals
	2	4	6	
+	39	41	43	123
0	7	12	9	28
−	8	7	2	17
Total	54	60	54	168
	$\chi^2 = 4.75$, $df = 4$			
	Duration of Participation			
	1 Hr.	2 Hrs.	3 Hrs.	
+	39	43	41	123
0	9	9	10	28
−	8	4	5	17
Total	56	56	56	168
	$\chi^2 = 1.95$, $df = 4$			

TABLE 4
COMPARISON OF MEAN SOCIAL INFLUENCE SCORES BY *t* TEST: IDENTIFIABLE VERSUS NONIDENTIFIABLE SOURCE OF RESPONSE STANDARD

Condition	Mean Social Influence Score	σM	*t*	*df*
Nonidentifiable source	4.00	.237		
Identifiable source (Two-member group)	6.00	.434	4.01*	30
Identifiable source	5.63	.254	5.14*	68

* $P < .01$.

Task Efficiency

To test the effect of group size on task effectiveness, performance measures were obtained on two of the three tasks assigned to the groups and individuals across conditions. The anagram task required differential instructions, so it was eliminated from consideration. The remaining tasks, sentence completion and jigsaw puzzle, were analyzed in two ways: with the whole group considered as the producing unit and with the individual member considered as the producing unit.

For the sentence completion task, efficiency was measured by sentences completed per unit time. No significant differences attributable to group size were observed when analysis was based on the group as the producing unit. Considering the individual as the producing unit, a significant *F* ratio of 57.78 was obtained ($p <$ .01). For the jigsaw puzzle task, completion time for the three puzzle solutions accomplished in a single session was pooled to give a mean completion score. As in the sentence completion task, when the data were analyzed

TABLE 5
ANALYSIS OF VARIANCE OF TASK EFFECTIVENESS OF VARIOUS SIZED GROUPS ON SCRAMBLED SENTENCE JIGSAW PUZZLE TASKS

Task	Source of Variance	*df*	Mean Square	*F*
Scrambled Sentence	Team Size	3	25.6	57.78*
	Within conditions error term	61	.443	
Jigsaw Puzzle	Team Size	3	661.0	21.05*
	Within conditions error term	61	31.4	

* Significant beyond the .01 level of confidence.

with the individual taken as the producing unit, the F ratio was 21.05, which is also significant beyond the .01 level. In both instances, the trend was one of decreasing task effectiveness as the number of co-acting group members increased. Table 5 is a summary of the analyses of variance performed using the individual member as the producing unit.

DISCUSSION

Social Influence

Contrary to the expectations expressed in the original hypothesis, group size was not shown to be effective over the range of two, four, and six members. This result confirms and extends the findings obtained by Goldberg (4) for two- and four-member groups. It is inconsistent, however, with Asch's results (1). All three studies nevertheless show the same general trend. Goldberg's *S*s in four-member groups were influenced more than their counterparts in two-member groups—but not significantly more. The finding of a consistent, though as yet nonsignificant, trend such as this should encourage investigators to continue attempts to isolate the effect of group size upon social influence phenomena. It is certainly possible that the lack of clearly positive results reflects the lack of precision in measuring instruments.

Again, contrary to the original hypothesis, duration of group participation was not revealed as an effective factor in social influence. This result confirms Bovard's finding (2) in a substantially disparate set of circumstances. The fact that in both studies all groups were relatively homogeneous with respect to age, college attendance, etc., is probably a relevant consideration. A conclusion that overreaches the data but seems reasonable is that among *S*s with well-established habits of peer-group interaction, the social influence response reaches an asymptotic value within a few moments of group contact.

With regard to the comparison of the easily identified as opposed to the relatively nonidentifiable source of the judgmental standard, there were measurably greater social influence effects when the source of the standard was known to be fellow group member(s). This is not to say that conformity did not occur when the source of the standard was anonymous. On the contrary, there was still a measurable tendency, in some *S*s at least, to modify their judgments in the conforming direction even when they have no ready referent for the source of the standard. But when this standard arises within a group of which they are even temporarily and superficially a part, the conformity response is considerably strengthened.

Task Efficiency

The data fail to confirm the hypothesis that efficiency would increase as size increased when the group was taken as the producing unit. This finding is inconsistent with the results obtained by Shaw (8) and by Taylor and Faust (10), but agrees with those of Moore and Anderson (7).

The contrary finding for two of the tasks that, with the individual taken as a producing unit, efficiency declines as group size increases tends to confirm the observations of both Taylor and Faust and of Moore and Anderson when they treated their data in individual rather than group terms. The finding appears to be rather consistent over a variety of tasks and under quite different experimental conditions.

Taking the group as the unit of production, however, leads to a marked disagreement. The Shaw and the Taylor and Faust studies found improvement with increased group size; in the Moore and Anderson and the present investigations, no differences were found between individuals and groups of various sizes. It is quite possible that the discrepancy is attributable to factors in the tasks or the general experimental operations. But conjectures can be advanced that resolve the discrepancy on theoretical grounds.

If one assumes that the group members can exert a restraining effect or a releasing effect on their fellow group members, the possibility exists that when the problem-solving activity is sustained over a considerable duration (as it was, for instance, in the Shaw study) a disproportionate effect is produced by the members with the greatest ability to solve the particular problem at hand. If the experiment is started with two samples equivalent in problem-solving ability, and one sample is divided into groups, the group members who are poor in problem-solving ability should affect the final outcome to a lesser degree than should persons of low ability who work on the

problems alone. Where groups have been found superior to individuals in task effectiveness, it is suggested that the outcome may be attributable to the fact that in the group situation some members were restrained from acting and others were encouraged to act—depending upon their apparent task ability.

SUMMARY

The effects of group size, duration of group participation, and identifiability of the source of a response standard were evaluated with respect to a measure of social influence. A total of 182 *Ss* were observed. A factorial design was used to test group size of two, four, and six members with duration of participation at 1, 2, and 3 hrs. An additional condition with single *Ss* was introduced to evaluate the effects of the identifiability of the source of the response standard. Cooperative group activity on three tasks was followed by a single measurement of social influence on a nontask-relevant psychophysical judgment.

For the particular situation studied, neither group size nor duration of group participation had significant effects on social influence. However, identifiability of the source of the response standard was accompanied by increased social influence.

A secondary area of investigation concerned the role of group size on task effectiveness. The results from a sentence completion and a jigsaw puzzle task were evaluated in this connection. With the group considered as the producing unit, increasing group size did not facilitate task performance. On the contrary, with the individual considered as the producing unit, increasing the number of co-acting members actually decreased the efficiency of task performance.

REFERENCES

1. ASCH, S. E. Some forms of interpersonal influence. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev.) New York: Holt, 1952. Pp. 2–11.
2. BOVARD, E. W. Conformity to social norms and attraction to the group. *Science,* 1953, **118,** 598–599.
3. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist,* 1955, **10,** 191–198.
4. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.,* 1954, **49,** 325–329.
5. HOVLAND, C. E., JANIS, I. L., & KELLY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
6. KIDD, J. S., & CAMPBELL, D. P. Conformity to groups as a function of group success. *J. abnorm. soc. Psychol.,* 1955, **51,** 390–393.
7. MOORE, O. K., & ANDERSON, S. B. Search behavior in individual and group problem solving. *Amer. soc. Rev.,* 1954, **19,** 702–714.
8. SHAW, M. E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.,* 1932, **XLIV,** 491–504.
9. SHERIF, M. *The psychology of social norms.* New York: Harper, 1936.
10. TAYLOR, D. W., & FAUST, W. L. Twenty questions: efficiency in problem solving as a function of size of the group. *J. exp. Psychol.,* 1952, **44,** 360–368.

Received December 3, 1956.

PERSONALITY AND PSYCHOPATHOLOGY[1]

E. I. BURDOCK, SAMUEL SUTTON, AND JOSEPH ZUBIN

Biometrics Research, New York State Department of Mental Hygiene and Columbia University, New York

THE hypothesis that the phenomena of mental disease are the outward manifestations of impaired brain function suggests a systematic exploration of the various dimensions of behavior to reveal the extent and the depth of defect. It rests with biometrics, the science which applies quantitative analysis to the structure and function of living organisms, to provide a map for such a survey by pointing out the main avenues along which progress may be expected, to integrate observations obtained at different levels of function, and to devise new tools for the extraction of significant information.

BIOMETRIC MODEL FOR ANALYSIS OF BEHAVIOR

An analysis of individual behavior can be organized by allocating observations among five major levels of activity: (*a*) physiological, (*b*) sensory, (*c*) perceptual, (*d*) psychomotor, and (*e*) conceptual.

Observations of such autonomic functions as heart rate, blood pressure, rate of respiration, production of sweat or saliva, etc., tap an area of behavior which can be measured without the overt cooperation of the subject (*S*) and even without his attention. In particular, while such measures are noticeably sensitive to external stimulation, they can nevertheless be obtained when the *S* is at rest, in what might be called the "idling" state, so that it would be possible to establish base lines of psychophysiological activity against which the influence of extrinsic stimuli could be more accurately assayed than at present.

Sensory activity refers to the modality-specific component of the complex of behavior involved in the functioning of a receptor organ. The immediate experience correlate of sensory activity in the human sensorium is a sensation, e.g., the mere awareness of light when the retina is illuminated. The neural conditions for a sensation are as follows:

1. Electromagnetic or gravitational energy impinging on a sense organ or free nerve ending.
2. Conduction of a nerve impulse by the specific afferent system from the receptor.
3. Collateral nonspecific afferent conduction in the reticular activating system (3, 31).
4. Efferent feedback regulation of the sensitivity level of the receptor (9, 11, 22).
5. Integration of peripheral, central, and autonomic neural activity at the level of the diencephalon (26).

The absence of the first condition, external energy, distinguishes a hallucination from a true sensation. One test that has been suggested for this distinction is the occurrence of after-images (41, pp. 6–9); but there is evidence that both eidetic images (17) and visual hallucinations (40) may be followed by after-images. Pure sensation is probably infrequent. Seeing light, feeling pain, hearing sound, etc., with no further qualification approximate the experience, but probably involve some conceptual abstraction on the part of the *S*.

In contrast to sensation, which is concerned with the unique quale of a sense modality, perception is an activity of the organism which superposes on the sensory function, scanning, focusing, and following mechanisms that are experienced as discrimination and that are oriented to reaction. In addition to the neural events involved in sensation, perceptual behavior incorporates primary efferent pathways into the circuit. Perception is thus the experience correlate of the integration of the primary afferent and efferent systems (30).

When overt movements accompany the perceptual act, the behavior is psychomotor. It is likely that some psychomotor activity accompanies most perceptual behavior.

The expression, conceptual behavior, is best reserved for those obscure components of the cortical network which mediate linguistic and metalinguistic activity.

Elsewhere (4) it has been pointed out that these five categories comprise an ascending hierarchy with physiological behavior providing the vegetative substrate and with conceptual behavior involving at least the first three

[1] Presented in part before Section I of the AAAS, New York, December, 1956. The preparation of this paper was facilitated in part by research grant, M-586 (C3) Psychological Prognosis in early and chronic schizophrenic patients, from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

levels as precursors or concomitants. A comprehensive account of an individual's mental functioning ought to include representative samples from each distinguishable level of behavior.

It is just as inadequate to evaluate psychopathology solely in terms of conceptual functioning as it would be in physical science to account for the phenomena of solar radiation only in terms of terrestrial combustion and without regard to the behavior of nuclear plasmas at extreme elevations of temperature. Indeed, the enormous energy expended by the sun without appreciable loss of mass remained a mystery until the submolecular aspects of the system were brought into meaningful relation with the molar observations by Bethe's hypothesis of the carbon cycle.

For psychopathology, biometrics has the responsibility of searching through the whole range of human behavior for the critical relations between different levels of function which will throw light on the mystery of mental disease. It is already abundantly clear from the history of psychopathology that those investigators who confine their attention to a single level of behavior cannot hope to account for the problem in all its complexity.

A map of the world of psychopathology needs meridians to distinguish the experimental techniques that can be applied across the latitudes of behavior. In studying the behavior of an individual, the experimenter may adopt one of three methods: First, he may merely observe the *S* without intruding in any way into the natural flow of behavior; second, he may seek to engage the *S* in some specific activity; or third, he may modify or distort the intrinsic functioning of the *S* and then proceed to examine the residue of behavior by one of the above two methods. The first procedure, naturalistic observation, is propaedeutic to experimentation. Observation and description provide a context from which hypotheses can emerge for testing under controlled conditions. Even the ethologists (27), who rightly insist on the importance of studying animal behavior in naturalistic settings, nevertheless do not hesitate to manipulate the environment so as to determine the limits within which a particular act, such as a duck's retrieving an egg, will be evoked.

The second method, the controlled experiment, has been the traditional method of choice of the experimental psychologist. A significant feature of experimental procedure is that the *S* is removed from his natural habitat to the artificial environment of the experimental laboratory.

In order to explore the intrinsic behavioral repertoire of an individual, the experimenter usually seeks to bring the *S* into reaction with some limited aspect of the environment which can be varied qualitatively or quantitatively. He then observes the activity of that segment of the organism which is peculiarly sensitive to the slice of the environmental surroundings that the experimenter is manipulating. The experimenter frequently establishes a baseline for his study by initiating his record before he has introduced the variable quality or quantity. The behavior of the organism corresponding to such a zero point may appropriately be described as in an "idling" state. With the organism in the idling state, the experimenter may conduct observations at any one of the five behavioral levels described above (physiological, sensory, perceptual, psychomotor, and conceptual.) Characteristically, the experimenter restricts his observations to one of these levels. Indeed, in this age of specialization, an entire professional lifetime may be devoted to the detailed study of a single behavioral level, or, more commonly, of some particular subdivision of such a level, e.g., auditory sensitivity (sensory), visual form discrimination (perceptual), word associations (conceptual), etc.

Just as the domain of individual behavior can be analyzed into five levels, so too can the parameters varied by the experimenter be differentiated into two broad classes above the zero point with its associated "idling state": energy variables and signal variables.

Energy variables are stimuli which furnish the necessary driving energy for a response. Characteristically, the stimulus sequence is graduated in units of intensity, duration or wave length (or its reciprocal, frequency). As the response is powered by the input energy, the experimenter seeks to express the corresponding activity in the organism as a mathematical function of energy. Energy variables are commonly used to study sensory functions, such as thresholds or j.n.d.'s. The energy may be appropriate, i.e., within the range to which

the sense organ is particularly responsive, as light to the eye, sound to the ear; or it may be *inappropriate*, i.e., outside the energy range for which the organ is specially adapted, as electrical stimulation above the retina to produce phosphenes, or caloric stimulation of the vestibular apparatus.

Signal variables, on the other hand, provide a relatively insignificant fraction of the energy released in the response. They are triggering devices which vary in spatial or temporal organization. Correspondingly, the response, deriving its motive power from energy intrinsic to the organism, can no longer be represented as a function of input energy.

The simplest kind of signal variable is the *configuration* in which the stimuli consist of spatio-temporal patterns or arrangements. Variations of juxtaposition or of time interval produce corresponding changes in behavior, as in the visual discrimination of shape or the auditory discrimination of timbre.

A more complex kind of signal variable is the *sign*. A sign is any event which implies some other event. More generally, a sign places a probability valuation on the event it designates. Thus, smoke is a sign of fire; a footprint in the sand signifies that someone has stepped there; a certain residue of C^{14} in a fossil plant is a sign of its age. Natural signs are either precursors or consequences of the events they designate. It is, however, a universal characteristic of human behavior to employ sets of arbitrary conventional signs for cultural purposes. For example, an arrow indicates direction; a red light warns of danger; a warbling tone on a siren gives the all clear. A set of signs may be said to comprise a code when arranged in some lexical order over against a corresponding set of designated events. The process of transmitting and receiving such encoded signs is the signal engineer's definition of communication. In the psychological laboratory, if the experimenter places an arbitrary sign (CS) in temporal or spatial contiguity with a reinforcing event (US), the sign by its presence or absence establishes a pattern, or configuration, which may be varied according to a predetermined sequence. Classical Pavlovian conditioning, animal experiments in visual or tactile discrimination, and delayed reward experiments employ this technique.

A *symbol* is a sign which designates an aggregate or class of events. Symbols are by their very definition abstract, removed from direct reference to particular events. Language, i.e., the systematic designation of classes of objects, attributes, or events by accidental patterns of acoustic energy, is in essence a complex communication code composed of symbols. In all likelihood, linguistic behavior is the prototype from which symbolism in art and music is derived. In origin, all music is "program music," i.e., has content reference, and all art is representative. The complex experience evoked by a work of art through the interplay of elements of pure design with subtly suggestive allusions to persons and events is probably prepotent in such devices as the Rorschach and the TAT. These techniques should therefore be viewed as presentations of symbol variables, and the corresponding behavior should be designated conceptual rather than perceptual. Other instances of behavior evoked by symbols are found in the interview and in the questionnaire. Table 1 shows examples of measurable activities as functions of stimulus variables.

The Modified Organism

On the premise that the whole is the sum of its parts, physiologists have long had the practice of studying isolated systems, e.g., the nerve-muscle preparation, the perfused liver, etc. Taking a leaf from this common physiological procedure, experimental psychologists and psychiatrists have in recent years begun to study the organism's behavior after first introducing certain modifications or distortions of intrinsic functioning. This has been done in three ways: imposition of a load, intervention in the internal neurohumoral environment, and ablation or destruction of part of the organism.

1. *Loads.* A load is defined as an externally imposed force which induces a distortion or strain and a counteracting stress. In a physical object, e.g., a wire with a weight attached, the restorative stress is proportional to the distortional strain, in accordance with Hooke's Law. Similarly, when a psychological load is imposed on an organism, the organism too undergoes a strain which is countered by restorative stress, although it is still moot whether an analogue of Hooke's Law can be established for psychology. Examples of psychological

TABLE 1

EXAMPLES OF MEASURABLE ACTIVITIES AS FUNCTIONS OF STIMULUS VARIABLES

Level of observed behavior	Stimulus Variables											
	Idling State		Energy Variables				Signal Variables					
	None		Appropriate Energy		Inappropriate Energy		Configurations		Signs		Symbols	
	Variable	Function	Variable	Function	Variable	Function	Variable	Function	Variable	Function	Variable	Function
Conceptual	—	Reverie and fantasy	Uniformly diffused light	Fantasy	Inversion of gravitational attraction	Subjective account of experience	Aircraft forms or silhouettes	Recognition of identity of forms	Practical trouble-shooting test	Diagnosis of trouble	Stimulus word in association test	Association to stimulus word
Psychomotor	—	Spontaneous movement	Painful stimulus	Arm withdrawn	Electroshock	Movements of limbs	Star-shaped maze	Mirror tracing	Classical delayed response stimuli in animal experimentation	Successful response by animal subject	Reinforcement of affect in focused interview	Electromyographic response
Perceptual	—	Spatial and temporal orientation	White noise	Orientation to direction of sound	Pressure stimulation above retina	Phosphene	Rotating Benham Disc	Subjective color experience	Usual visual alternatives in animal discrimination experiment	Selective response of animal subject	Musical chords	Pitch discrimination
Sensory	—	Background noises; cortical gray	Light of graded intensity	Threshold response	Electrical stimulation of thermal receptors	Warmth or cold sensation	Patterned light stimuli	Visual threshold	Infant's faint cry	Mother's auditory threshold	Words or sentences presented tachistoscopically	Visual threshold
Physiological	—	BMR; basal EEG; basal PGR	Increase in carbon dioxide concentration	Change in rate of respiration	Pressure on carotid sinus	Change in heart rate	Photic driving	Change in EEG pattern	Bell ringing in Pavlovian conditioning	Salivation	Verbal instructions to prevaricate	Effect on PGR

loads are threat, with consequent strain of anger or fear (5); tissue damage, with the strain reflected in the inflammatory process; deprivation of air, food, water, sensory input, etc. (2, 15); and intense physical effort or prolonged exposure to extremes of temperature. Just as an excessively severe physical load will produce a permanent distortion in a wire, so too can a psychological load strain an organism beyond its elastic limit, so that subsequent strains must be measured from a new baseline. The consequences of psychological loads which are excessively severe or unduly prolonged are seen, for example, in "combat fatigue," in the General Adaptation Syndrome (39), or in the adrenal hypertrophy which follows intense uninterrupted physical exertion (35).

2. *Neurohumoral intervention.* The application of drugs or brief electrical current to produce activity in the organism has become increasingly popular with experimenters. Drugs like mescaline (18), adrenochrome (14), and LSD (42) have been used to evoke bizarre behavior or to initiate anomalies of subjective experience in the hope of uncovering a biochemical key to psychopathology. Despite many suggestive analogies, however, their effects are obscured by the wealth of interactions which they set up in the organism. The difficulties to be surmounted before a biochemical understanding of psychopathology can be expected may be appreciated when it is realized that the biochemistry of so common an agent as aspirin is still obscure. The initiation of behavior by intracranial electrical stimulation, first used for mapping the brain (8), has recently become a method of choice for the production of activity or affect (6, 32). But here again the relevance of these new techniques for the behavior of the intact organism remains to be demonstrated.

3. *Ablation.* The method of ablation, by which a part of the brain substance or of the neurohumoral network is excised, has made many substantial contributions to the understanding of the structure and function of the nervous system (1, 19, 38). However, the caution cannot be repeated too often that what is subsequently observed is the residual behavior of all the rest of the animal, including whatever compensatory or suppressed mechanisms may have been released. Inferences as to the role of the excised part in behavior remain at best dubious. This is particularly true because of the difficulty of assessing the side effects of trauma, degenerative changes, bleeding, anesthesia, etc.

Of the three methods for modifying the organism, the imposition of a load seems at present the most promising for psychopathology because it usually brings about a response that differentiates between mental patients and normals. Neurotics and early psychotics tend to react with excessive strain, while chronic schizophrenics show a deficiency in reactivity. In studying behavior under load, some care must be exercised to avoid confusing the load with the independent variable. A load which causes tissue damage characteristically takes the form of an energy variable. But a psychological threat may present itself as a sign or a symbol. Particularly when one is studying behavior of the "idling" state under load, there is a tendency, because of the absence of an independent variable, to think of the load as the variable. However, if the term *load* is restricted to those environmental factors which menace the integrity or existence of the organism, the distinction will serve a useful purpose.

INTEGRATION OF OBSERVATIONS

It has been noted for some time that the usual performance of the mental patient may not differ from the performance of normals, provided attention and motivation are sustained. It is quite possible that in chronic schizophrenics the prolonged strain induced by the load which the disease itself constitutes has brought about a permanent distortion in the patient's adaptive capacity so that he displays reduced sensitivity under any additional load. It is, therefore, important that autonomic base lines be established for patients vis-à-vis normals under the idling state and that other variables be studied both from the point of view of usual functioning and from the point of view of functioning under a load. The hypersensitivity displayed by early schizophrenics under load in contrast to the apathy of the chronics has already been noted. It may very well be that no single aspect of behavior will alone distinguish the patient from the normal, but that the pattern of function will be distinctive (23). An additional point that has been noted (47) is that patients whose conceptual functioning is depressed below their

perceptual function are more likely to improve than those with conceptual behavior elevated above perceptual.

TOOLS AND TECHNIQUES OF RESEARCH

A program designed for the intensive exploitation of these new leads in psychopathology is represented in Table 2 in which certain techniques that have shown promise in preliminary undertakings are classified according to the schema previously described. The omission of techniques for the study of sensory activity is justified by the fact that sensory thresholds have not been found to bear any relation to outcome of mental illness. On the other hand, it should be emphasized that techniques which have failed to discriminate patients from normals under ordinary conditions may yield significant differences when the subjects are tested under a load, such as psychological threat, sensory deprivation, or extreme of temperature. The imposition of a load reveals the lowered tolerance and decreased resilience of the patient.

Physiological Techniques

The increasing evidence that schizophrenics may differ from normals in essential physiological function makes it urgent that experimental psychopathology give greater attention to the physiological level of behavior. The following two techniques represent efforts to measure autonomic components of physiological activity.

1. *Pupillography.* The method of recording pupillary reflexes developed by Lowenstein (28) presents an unusual opportunity to study the interaction between the sympathetic and parasympathetic divisions of the autonomic system, since presumably both divisions modulate the pupillary mechanisms. In this technique, serial photographs of the eyes are made under infrared illumination after dark adaptation and during controlled exposures to light. The diameter of the pupil is measured from the film and plotted against time. Lowenstein and his co-workers have performed numerous animal experiments and have assembled a large body of clinical data which point to relationships between pupillographic curves and specific lesions or dysfunctions of the central nervous system. This technique was applied to the study of schizophrenia in a pilot investigation by Lowenstein and Westphal in 1933 before the method had been satisfactorily standardized. The curves they obtained for catatonic schizophrenics were strikingly different from those of normals. The diminution of the pupillary reflex on repeated evocation is more pronounced and appears earlier in patients. Moreover, in normals, but not in schizophrenic patients, the initial level of response can be restored by inducing startle, i.e., when a sudden intense sound is introduced as a load, patients react differently from normals. It has also been noted that the pupillograms of schizophrenic patients show striking similarities to those of patients with extrapyramidal lesions.

2. *Pattern of autonomic responsivity.* Many investigators have reported differences between schizophrenics and normals in autonomic responsivity, particularly under conditions of stress-producing load. There is some indication that acute cases tend to overreact, chronic cases to underreact. However, the equivocal results of a substantial number of studies point up the need for further investigation. It is clear that the degree of activation of autonomic mechanisms is correlated with the strength of the stimulus, or load, ranging from simple alerting responses all the way to the severe and prolonged reactions evoked by combat fatigue. Research by Lacey (23) and by others has also shown that an individual's pattern of autonomic responses may be invariant despite great differences among individuals.

Taken together, these findings suggest the need to sample patterns of autonomic responsivity at several points along the intensity dimension. It may be that failure to find differences between schizophrenics and normals on a single response variable is caused by individuals with contrasting autonomic patterns who cancel out one another's responses in both groups. A polygraphic study of schizophrenics and normals provides a technique for determining the autonomic pattern as a basis of comparison. Such a study should include techniques from each class of stimulus variables, beginning with the idling state as a baseline and progressing through the energy operators and the signal operators. In particular, it is to be observed that the imposition of the load constituted by sensory deprivation may be expected to magnify the differences to be found between patients and normals.

TABLE 2

CLASSIFICATION OF TECHNIQUES FOR STUDY OF EARLY AND CHRONIC SCHIZOPHRENICS

Technique	Variable		Function		Load	Reaction
	Class	Description	Class	Description		
Pupillography adaptation and restitution	appropriate	alternating light and dark	autonomic	contraction and dilation	sudden intense sound	restitution
Pattern of autonomic responsivity	none (idling state) appropriate configural sign symbol	— flash of light to eye Archimedes Spiral colored light as CS interviewer's activity in affect-focused interview	autonomic	heart rate, blood pressure, P.G.R., respiration rate, etc.	sensory deprivation	corresponding changes in functions
Weight judgment unanchored anchored	configural configural	weight weight	perceptual perceptual	discrimination of wts. discrimination of wts.	— —	
Flicker-fusion	configural	intermittent light	perceptual	discrimination of "steady" from "flicker"	—	
Necker cube	configural	ambiguous figure	perceptual	discrimination of alternative figures	—	
Archimedes spiral rotation end of rotation	configural configural	visual pattern visual pattern	perceptual perceptual	perception of movement reversal of movement	— —	
Digital stimulator	appropriate	tactual stimulation of more than one finger	perceptual	discrimination between stimulated fingers	cold pressor	change in discriminative accuracy
Color-form test	configural	visual pattern	perceptual	report of direction of apparent movement	—	
Purdue pegboard	configural	pattern of pegs and holes	psychomotor	placement of peg	—	
Tapping reaction time tapping	sign configural	colored light pattern of tapping targets	psychomotor psychomotor	movement speed of motion	— —	
Reaction time shift	sign	visual and auditory	psychomotor	movement	threat	difference in latency
Delayed auditory feedback in reading	symbol	previous speech segment	psychomotor	articulation movements	pin prick	change in rate of articulation

Metalog	symbol	word	conceptual	definition	—	
Analogies	symbol	word	conceptual	eduction of correct analogue	—	
Essential differences	symbol	word	conceptual	discrimination of "different" word	—	
Porteus Maze	configural	visual pattern	conceptual	choosing direction	—	
Stroop I	symbol	word	conceptual	reading	shock	difference in latency
II	sign	color	conceptual	naming		
III	symbol	color word printed in conflicting color	conceptual	naming		
IV	symbol	color word printed in conflicting color	conceptual	reading		
Vocabulary	symbol	verbal utterance	conceptual	definition	—	
Affect-focused interview	symbol	interviewer's behavior	conceptual	affective utterance		—

Because of the critical significance of emotional changes in mental illness, it seems important that base lines of autonomic response patterning should be developed for the various psychiatric groups. Malmo (29) has recently studied pathological anxiety in this way. It may be possible to develop indices which will constitute objective correlates of global clinical concepts such as "flattened affect," "degree of tension," "process and reactive schizophrenia," etc. Such an achievement would be a considerable contribution to the prediction of outcome in schizophrenia and might aid in distinguishing clinical subgroups within the psychosis. By studying indices of autonomic reactivity which reflect the readiness of the organism to adjust to change, one may gain insight into some of the determinants of the duration of illness.

Perceptual Techniques

The significance of perceptual behavior for prognosis becomes more apparent when the overlay of conceptual activity is peeled off. Moreover, techniques which minimize the conceptual component, permit a comparison between the patient's perceptual and conceptual function. As intimated above, it may be the difference between these levels rather than their absolute function which distinguishes those who get well from the chronics.

3. *Weight judgment* (*method of absolute judgment*). This is a test of perceptual behavior in response to configural stimulation of varying complexity. The patient is given five weights to judge according to a range of categories from very heavy to very light. After the accuracy of judgment has been determined, a heavy anchor is introduced just before the presentation of each standard weight. The introduction of the heavy anchor tends to make the succeeding standard weight appear lighter. The *S* is instructed to judge only the second or standard weight. In this way, the directions attempt to offset the influence of the anchor. The effectiveness of the instructions is measured by the amount of shift in the judgment categories following the introduction of the anchor.

In a recent study, Salzinger (36) found that patients did not differ significantly from normals in accuracy of weight judgments, but did tend to shift significantly more than the normals, i.e., the schizophrenics were less successful in resisting the influence of the heavy anchor.

4. *Critical flicker frequency.* The power of the visual system to resolve stimuli closely spaced in time is a stable measure for the individual under basal conditions but is altered by damage to the central nervous system, by drugs, by illness, and by the normal process of aging (20, 24). There is some preliminary evidence that outcome of illness for schizophrenic patients may be related to level of cff at time of admission. In two studies, patients who were eventually discharged had *lower* thresholds at admission than those who remained in the hospital. As anxiety serves to reduce cff (21), it is a tenable hypothesis that patients who show greater anxiety, and hence lower cff, at time of admission have a better prognosis.

5. *Necker cube.* The number of fluctuations in the perception of an ambiguous figure, such as the Necker cube, can be counted by having the *S* press a switch every time he sees the cube shift. The data may then be analyzed for the number of shifts in each minute as well as for the total number of shifts.

Eysenck (7), following the lead of McDougall and of Hunt and Guilford, found that schizophrenics had a lower rate of reversals than normals. It is not surprising, then, as a recent study at Brooklyn State Hospital has disclosed, that first admission mental patients who twelve months later were out of the hospital had a higher rate of shift than those who remained in the hospital. This relationship, however, was reversed in a readmission group—with the "Ins" having a higher number of shifts than the "Outs." All together, the first admission Ins had the lowest number of shifts, and the readmission Ins had the highest number of shifts. Moreover, sicker patients reached a maximum rate of reversals earlier in the five minute observation period.

6. *Archimedes Spiral.* A rotating Archimedes Spiral will be seen as a contracting or expanding configuration, depending on the direction of rotation. When the rotation is stopped, a contrary after-effect is induced in normals, but not in some brain-injured patients (34). This technique is being introduced with the expectation that it may help to detect brain-injured cases among the functionally ill.

7. *Digital stimulator.* It has been known for some time that perceptual defects may be uncovered by use of simultaneous stimulation methods (13). Although little work has been done on schizophrenic patients, reports thus far available indicate that schizophrenic children show defects in tactual perception with the Face-Hand test. Adult schizophrenics also show perceptual defects when tested by simultaneous stimulation methods. A test apparatus has been designed to permit simultaneous stimulation of one to eight fingers at a constant intensity of stimulation. Preliminary analysis indicates that the amount and pattern of errors made by schizophrenic patients are different from those of matched controls.

8. *Color form test.* Thurstone's color-form test was developed for the objective measurement of color versus form preference. It consists of a moving-picture film in which the *S* indicates the direction of the apparent movement noted in the presented figure. If he sees the movement in one direction, he is color-dominant, if in the other, he is form-dominant. Some preliminary work with this technique indicates that depressed schizophrenics prefer the form responses and the downward-tending direction more often than do nondepressed patients.

Psychomotor Techniques

It has long been noted that mental patients differ from normals in their motor behavior. Indeed, the general apathy and sluggishness of many mental patients is one of their most striking characteristics, while at the other extreme the intense motor activity of the manic is often the first sign of abnormality. Recently, considerable evidence has accumulated (16) that psychomotor behavior may serve to distinguish subgroups among patients. It is in the hope of finding discriminators that will separate those who will eventually get well from the patients who fail to improve, that techniques 9 through 11 have been introduced.

9. *Purdue pegboard.* This is a standard test (44) of finger dexterity with a great deal of data on normals available for comparison. In recent studies, chronic patients performed more poorly than new admissions. Among the new admissions, significant differences were found between patients who were being hospitalized for the first time and those who had records of previous hospital admissions for mental illness.

10. *Tapping test.* A large number of studies have indicated a general slowing and an increased variability in choice reaction time in schizophrenic patients (16). Recent work has

tended to confirm these results and in addition patients with a good prognosis have been found to improve with successive testing, while those with poor prognosis showed no consistent trend. On the other hand, the trends reversed on this test for first admission patients and readmission patients. First admissions who performed well had a good prognosis, but in the readmission group poor performance was associated with good outcome. In general, the same trends hold for total tapping time as for choice reaction time, although the test seems to be somewhat less sensitive in distinguishing between groups.

11. *Reaction time for sensory modality shift.* This task requires the *S* to make a choice reaction to each of four stimuli, two in the visual modality and two in the auditory modality. These four stimuli are presented in a random series and the reaction times are classified into two groups: those for which the stimulus to the previous reaction is in the same sensory modality as the reaction under scrutiny; and those for which the stimulus to the previous reaction is in the other sensory modality. This technique is still in the pilot stage, but it is expected that schizophrenics will have a greater lengthening in reaction time than normals when there has been a shift from one modality of stimulation to the other.

Conceptual Techniques

The domain of conceptual behavior has for so long been the favorite habitat of the clinician that it hardly requires any justification, for it is indeed precisely at this level of functioning that mental disease manifests itself. Perhaps as a natural consequence, a confusion between categorization and description has vitiated much of the research at this level. In order to derive reliable evidence of conceptual functioning both for diagnosis and, more importantly, for prognosis, the psychopathologist must turn from the easy way of global observation up the stony path of rigorous measurement. Techniques numbered 12 through 19 represent some familiar and some new attempts to introduce quantification into the study of conceptual behavior. Here again the imposition of a stress-producing load helps to magnify the differences between patients and normals.

12. *Delayed auditory feedback.* The alterations in speech resulting from an experimentally induced delay in auditory feedback have been extensively studied. Individual reactions to delayed feedback vary from almost no disturbance to complete blocking of speech. The extent of the interference may represent the degree to which the individual makes use of air-conducted auditory feedback to steer ongoing speech. Experimental variation of the input to this steering mechanism provides a unique opportunity to study such processes in psychiatric populations. In a pilot study, some schizophrenics showed less disturbance than normals under delayed auditory feedback. These results corresponded with a preliminary study by Goldfarb (10) on schizophrenic children. Recently, normative and developmental studies were begun to obtain baselines for studies of abnormal populations. Preliminary results indicate substantial differences between normal children aged 4 to 6 and normal children in the 7- to 9-year age range.

13. *Metalog.* The Metalog test (46) attempts to tap facility of conceptual shift by requiring the *S* to produce as many different meanings to a simple stimulus word as he can within one minute. A high score corresponds to a large number of different meanings. This test was first used in connection with the Columbia-Greystone (25) study where a high score was found to be associated with poor prognosis in both the operated and unoperated groups.

14. *Analogies test.* In this test of conceptual performance, the *S* is required to complete a series of verbal analogies of the type, "Night is to day as dark is to *light*." Like the Metalog test, this test too was found to be inversely related to outcome for patients in the Columbia-Greystone study (25).

15. *Essential differences test.* In this test, the *S* is presented with successive sets of four words and is required to indicate in each set which word is most different from the other three; for example, apple, pear, *carrot*, plum. The *S* thus has to abstract an attribute which is common to three of the four words (25). He is also asked to justify his selection, i.e., to give an account of the ideational process involved. This is one of several tests which seem to be positively related to outcome for first admissions, but negatively related to outcome for readmissions.

16. *Porteus maze.* A standard test of nonverbal conceptual functioning, the Porteus (33) invokes foresight and planning. This test has shown significant changes after brain surgery.

Chronic schizophrenic patients tend to score quite poorly on this test.

17. *Stroop test.* This test (43) compares the ability to perform a conceptual task involving discrimination of four colors with and without distraction. The *S* is first asked to read a list of color names. His next task is to assign the proper names to a random sequence of colored dots. Then he is presented with a list of color names printed in a conflicting color of ink. He must name as rapidly as possible the color of the ink in which the words are printed. Finally, he is told to ignore the actual colors and to read the color names as fast as he can. The difference between the speed with which colors are named in the conflict situation and the speed when no conflict is present provides a critical score. Ability to resist the distraction of the conflict card appears to be positively related to outcome of schizophrenic illness.

18. *Vocabulary test.* Vocabulary can be considered a good indicator of overlearned conceptual behavior for purposes of comparison with other measures of conceptual functioning. The word list is taken from the Wechsler and Binet tests.

19. *Affect-focused interview.* Distortion of affective experience is one of the major characteristics of mental disorder. In fact, patients seldom appeal for help because of any inadequacy in adjustment to the environment. They usually come to the physician because of the way they feel. Moreover, one of the major findings of a survey of the literature was that patients with flattened affect had poor prognoses. In the affect-focused interview (12, 37, 45), the patient is first engaged in conversation for ten minutes (the operant period). During the second ten-minute interval, every affective utterance is reinforced by expression of assent, such as "umhum" (conditioning period). In a third ten-minute interval, the reinforcement is withdrawn (extinction period). An analysis of recorded interviews for the number of affective utterances during the operant, conditioning, and extinction periods gives a measure of the ability of the patient to emit affective utterances under reinforcement. The procedure has been refined by controls for reliability, i.e., repeated interviews with different interviewers, and repeated interviews on different occasions with the same interviewer. The interview is conducted in a room provided with a one-way observation screen so that the nonverbal behavior of interviewer and interviewee can be observed.

SUMMARY

The aid of biometrics is invoked for the design of a program of experimentation; for the integration of findings at different levels of complexity; and for the development of new research tools.

The behavior of individuals can be analyzed into a set of overlapping domains of ascending psychological complexity: physiological, sensory, perceptual, psychomotor, and conceptual. These areas have not received equal attention from students of psychopathology. In the past, there has been too much preoccupation with the conceptual domain. In fact, a good deal of research labeled as perceptual is actually concerned with conceptual behavior. While sensory psychology has thus far contributed little to our understanding of psychopathology, the neglect of physiology by most psychopathologists is unfortunate. Moreover, the interrelations of these functional levels may prove to be of prognostic value. Thus, there is already some evidence that the ratio of perceptual to conceptual function distinguishes different subcategories of mental patients.

The variety of investigative methods employed in research ranges from naturalistic observation through controlled experiments to modifications of the organism. The first method, observation, is propaedeutic to science. The laboratory experiment involves directing energy or signals at the organism and studying that part of the subsequent behavior which can be functionally related to the independent variable. Variables of the energy category may be appropriate (e.g., light for the eye) or inappropriate (e.g., pressure on the eyeball). Signal variables can be: simple configurations (as in studies of perceptual discrimination); signs which represent some event remote in time or space (as in learning experiments); or symbols (as in studies of conceptual function.)

In order to throw additional light on the behavior of the intact organism, it is often useful to study response to controlled stimulation when the subject is under the influence of a stress-producing load. By straining the organism to its limit of resilience, a load may help to reveal differences between normals and

patients which are obscured under everyday conditions.

Radical invasions of the organism's integrity, such as administration of hallucinogenic drugs, implantation of electrodes or ablation of brain tissue, can provide information about limited subsystems of behavior, but raise many problems of relevance with respect to integrated behavior.

A comprehensive program of research is presented, part of it already under way, designed to exploit the possibilities inherent in the interactions between different levels of psychological function.

REFERENCES

1. BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behaviour. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 362–404.
2. BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Can. J. Psychol.*, 1954, **8**, 70–76.
3. BREMER, F. Dualité des processus d'excitation centrale. *Ann. Physiol. Physiochim. biol.*, 1933, **9**, 897–903.
4. BURDOCK, E. I., & ZUBIN, J. A rationale for the classification of experimental techniques in abnormal psychology. *J. gen. Psychol.*, 1956, **55**, 35–49.
5. CANNON, W. B. *Bodily changes in pain, hunger, fear, and rage.* New York: Appleton-Century-Crofts, 1939.
6. DELGADO, J. M. R., HAMLIN, H., & CHAPMAN, W. P. Technique of intracranial electrode implacement for recording and stimulation and its possible therapeutic value in psychotic patients. *Confin. neurol. (Basel)*, 1952, **12**, 315–319.
7. EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality. *J. Pers.*, 1952, **20**, 345.
8. FRITSCH, G., & HITZIG, E. Ueber die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Physiol. wiss. Med.*, 1870, **37**, 300–332.
9. GALAMBOS, R. Neural mechanisms of audition. *Physiol. Rev.*, 1954, **34**, 497–528.
10. GOLDFARB, W. Reactions to delayed auditory feedback among a group of children institutionalized for severe behavior disorders. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of communication.* New York: Grune & Stratton, in press.
11. GRANIT, R. Centrifugal and antidromic effects on ganglion cells of retina. *J. Neurophysiol.*, 1955, **18**, 388–411.
12. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
13. HEAD, H., & HOLMES, G. Sensory disturbances from cerebral lesions. *Brain*, 1911, **34**, 102–254.
14. HOFFER, A., OSMOND, H., & SMYTHIES, J. Schizophrenia: A new approach. II. Result of a year's research. *J. ment. Sci.*, 1954, **100**, 29–45.
15. KEYS, A., BROZEK, J., HENSCHEL, A., MICKELSEN, O., & TAYLOR, H. L. *The biology of human starvation.* 2 vols. Minneapolis: Univer. of Minnesota Press, 1950.
16. KING, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954.
17. KLÜVER, H. An experimental study of the eidetic type. *Genet. Psychol. Monog.*, 1926, **1** (No. 2), 69–230.
18. KLÜVER, H. *Mescal: The divine plant and its psychological effects.* London: K. Paul, Trench & Trubner, 1938.
19. KLÜVER, H., & BUCY, P. C. Preliminary analyais of functions of temporal lobes in monkeys. *Arch. Neurol. Psychiat.*, 1939, **42**, 979–1000.
20. KRASNO, L. R., & IVY, A. C. The response of the flicker fusion threshold to nitroglycerin and its potential value in the diagnosis, prognosis, and therapy of subclinical and clinical cardiovascular disease. *Circulation*, 1950, **1** (No. 6), 1267–1276.
21. KRUGMAN, H. E. Flicker fusion frequency as a function of anxiety reaction: An exploratory study. *Psychosom. Med.*, 1947, **9**, 269–272.
22. KUFFLER, S. W., & HUNT, C. C. The mammalian small nerve fibers: a system for efferent nervous regulation of muscle spindle discharge. *Assn. Res. Nerv. Ment. Diseases*, Research publ. **30**, 1952, 24–47.
23. LACEY, J. I., BATEMAN, D. C., & VAN LEHN, R. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15**, 8–21.
24. LANDIS, C. Determinants of the critical flicker fusion threshold. *Physiol. Rev.*, 1954, **34** (No. 2), 259–286.
25. LANDIS, C., GARRISON, M., HAMWI, VIOLET, STAUFFER, ANNE K., KING, H. E., & YOUNG, KATHLEEN M. Test results one year after operation. In F. A. Mettler (Ed.), *Selective partial ablation of the frontal cortex.* New York: Paul B. Hoeber, 1949. Pp. 302–305.
26. LASHLEY, K. S. Dynamic processes in perception. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 422–443.
27. LEHRMAN, D. S. A critique of Konrad Lorenz's theory of instinctive behavior. *Quart. Rev. Biol.*, 1953, **28**, 337–363.
28. LOWENSTEIN, O. Pupillography. *AMA Arch. Ophthal.*, 1956, **55**, 565.
29. MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24.
30. MEAD, G. H. *Mind, self and society.* Chicago: Univer. of Chicago Press, 1934.
31. MORISON, R. S., DEMPSEY, E. W., & MORISON, B. R. Cortical responses from electrical stimulation of the brain stem. *Amer. J. Physiol.*, 1941, **131**, 732–743.
32. OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 1954, **47**, 419–427.
33. PORTEUS, S. D., & PETERS, H. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.*, 1947, **36**, 3–86.

34. PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
35. RICHTER, C. P. Total self regulatory functions in animals and human beings. *Harvey Lectures*, 1942–43, **38**, 63–103.
36. SALZINGER, K. Shift in judgment of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. Unpublished doctoral dissertation, Columbia Univer., 1954.
37. SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. Paper read at East. Psychol. Ass., New York, April, 1957.
38. SCHREINER, L., & KLING, A. Behavioral changes following rhinencephalic injury in the cat. *J. Neurophysiol.*, 1953, **16**, 643–659.
39. SELYE, H. The general adaptation syndrome and the diseases of adaptation. *J. clin. Endocrinol.*, 1946, **6**, 117–230.
40. SIDGWICK, H. *et al.* Physiology of hallucinations. *Proc. Soc. psychical Res. (E.)*, 1894, **10**, 134–148.
41. SMYTHIES, J. R. *Analysis of perception*. London: Routledge & Kegan Paul, 1956.
42. STOLL, W. A. Lysergsäure-diäthylamid ein Phantasticum aus der Mutterkorngruppe. *Schweiz. Arch. Neurol. Psychiat.*, 1947, **60**, 1.
43. THURSTONE, L. L., & MELLINGER, J. J. *The Stroop test*. University of North Carolina, Psychometric Laboratory, No. 3, 1953.
44. TIFFIN, G., & ASHER, E. J. The Purdue pegboard: Norms and studies of reliability and validity. *J. appl. Psychol.*, 1948, **32**, 234.
45. VERPLANCK, W. S. The control of the content of conversation by reinforcement. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.
46. ZUBIN, J., & WINDLE, C. The prognostic value of the Metenym test in a follow-up study of psychosurgery patients and their controls. *J. clin. Psychol.*, 1951, **7**, 221–223.
47. ZUBIN, J., WINDLE, C., & HAMWI, VIOLET. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

Received January 30, 1957.

THE PSYCHODYNAMICS OF SOCIAL MOBILITY IN ADOLESCENT BOYS

ELIZABETH DOUVAN AND JOSEPH ADELSON

Survey Research Center, University of Michigan, and University of Michigan

THERE is a large and growing body of literature on social mobility, and the significance of the topic is generally acknowledged; so it is surprising to find that only limited attention has been given to studying the motivational sources of mobility. What we do find is a general disposition to treat *upward* mobility in a vaguely invidious fashion. It would seem that, in this country, the Horatio Alger tradition and the "dream of success" motif (6) have been pervasive and distasteful enough to have alienated, among others, a good many social scientists. The upwardly aspiring individual has apparently become associated with the pathetic seeker after success or with the ruthless tycoon. This image of success is, much of it, implicit—assumption and attitude, and not quite conviction—but it seems to have dominated the thinking of our intellectual community.

Or so it has been until recently. Newer empirical findings have encouraged a more differentiated view of mobility. We begin to get some sense that varying motivations may underlie social striving; we become aware that the direction, the rapidity, and the absolute extent of mobility may reflect differing psychodynamic sources and thus require separate analytic treatment.

In this paper, a first effort is made to explore and clarify one problem in this complex area—to analyze some of the psychic accompaniments of upward and downward mobility strivings among adolescent boys. In the following section, a theory of the personality determinants of mobility is presented; the paper as a whole is devoted to testing some of these formulations.

THEORETICAL CONSIDERATIONS

Upward Mobility

We can distinguish at least three separate patterns of motivation which are implicit in discussions of upward social mobility: fear of failure, ambivalence toward success, and hope of success.

The first pattern—perhaps the most common image of the mobile individual—is of the desperately scrambling Philistine, exhausted in his pursuit of status. He has defined his identity and personal worth exclusively by the criteria of success. Perhaps a typical genetic source of this pattern has been described by Ackerman and Jahoda (1). Competitiveness is taught by the family: "Most of the mothers of our cases . . . apparently did not tell their children 'be happy' but rather 'make money'. . . ." "Success is measured by comparison with others rather than by actual achievement. . . . There are always some who have done better, who have more money and more social prestige; and there is always the danger of being pushed down the social ladder by a competitor" (1, pp. 88–89).

In the second pattern—ambivalence toward success—the dynamics are more complex. The individual aspiring upward is here seen as responding to the dominance and prestige of the more privileged stratum through a defensive identification with this group and by denying and decrying his own background and status. The genetic sources of this choice of mechanism may be sought in the aspirer's early encounters with authority and status in the original family setting. We discover in the family scene a harsh and forbidding parent, one who allows the child no hostility; we find a child who has accommodated to parental strength through identification. In this formulation, we see the mobile individual showing many of the characteristics of the "authoritarian personality": a conscious over-idealization of the parent and of authority together with unconscious rebellion and hostility, rigidity, conformity, and anti-intraception. As regards mobility, we find the following dynamics: the motive for upward aspiration arises out of the need to emulate the parents; at the same time mobility arouses conflict since it implies a struggle with this powerful authority figure (2).

Without question, both of these mobility types occur (perhaps frequently) in our society. Yet the recent literature on mobility has suggested that the most common pattern is the

one we have called "hope of success" (7). Here we find interest in success directed by a rational ego: the individual who can mobilize his energies effectively, whose aspirations are realistic. The mobility goals are moderate or, in those cases where aspiration is toward a substantial status move, they are nevertheless realistic in view of the individual's talents.

We find a family milieu which encourages autonomy, is not obsessed by status, yet accepts and transmits without ambivalence the culturally central value of achievement. Moreover, the shifts in identity required do not implicate the child in conflict: the family has provided or permitted the child a model for identity consistent with the goal of mobility.

In a society which holds upward mobility to be a central value, which provides opportunity for it, and in which it is, apparently, a very common occurrence, we can expect that the dominant motivational pattern informing upward aspiration is not a defensive one—does not necessitate a personally damaging flight from one's past; rather, we would expect it to accompany effective ego functioning and successful socialization by a family which is transmitting a cultural value of which it approves.[1]

Downward Mobility

What about the adolescent boy whose aspirations are downward in direction? Here we have someone whose goals are atypical; indeed, they contradict directly a strongly held cultural value. To be sure, the same novelistic tradition which associates upward mobility with Babbittry has sometimes offered us a sentimental image of the downward mobile type, presenting him as an unfettered, romantically sullen child of nature. Our own view is considerably darker.

We see downward aspiration as representing, in many areas, the psychodynamic opposite of upward mobility. If we are correct, these adolescents are demoralized, alienated, anomic. An ambivalent relationship to the parents produces an impoverished ego, vulnerability to conflict, a failure to internalize general cultural values. Ambivalence is, of course, an almost universal outcome of the socialization process; but we conjecture that it is especially acute among the downward mobile. We expect to find a relationship to the parents which is infantile, that is, dependent and hostile. Dependency and aggression accompany, reinforce, and conceal each other. The child who is captured in an infantile tie to his parents has too much of his energy committed in ambivalence; ego functions remain immature and conflict-ridden; the ego is unable to cope adequately with impulses, internalized morality, and the more complex aspects of reality. At its best, the ego enforces a harmony among the psychic institutions and the outside world. But the downward mobile boy, if our formulation is correct, is at the mercy of his impulses, which continually threaten to overwhelm brittle controls. The superego drives are equally primitive and frightening; parental prohibitions are not sufficiently internalized, in part because of their very intensity; superego functioning is consequently immature, in that there is an inadequate fusion of ego and superego processes; the superego stands at some distance from the ego, so that questions of morality are not so much matters of right and wrong as of escape and pursuit. As regards the ego's operations vis-a-vis the outer world, we do not hypothesize that the downward mobile show any gross pathology of reality testing. And yet we feel that the more complex and articulated ego processes—judgment, rapport, control, time-binding—are poorly or unevenly developed.

Hypotheses

In the research reported here, a study of the mobility aspirations of adolescent boys, we

[1] One may speculate about those social situations in our culture where upward aspiration necessitates a good deal of emotional stress or induces personal conflict. Generally, we would say, these are situations where a change in status requires a decisive modification of behavior or necessitates a sharp shift in identity models. A number of such situations come to mind: (*a*) where the status change involves a concomitant ethnic (including religious) shift; (*b*) where it involves a gross movement up the status ladder; (*c*) where there has been a limited opportunity to learn the status behavior of the higher status group; (*d*) at certain points in the status continuum, where the defining criteria of the higher status group are realistically difficult to achieve; e.g., in the movement from lower-upper to upper-upper status, in those cases where a defining criterion is ancestry; (*e*) where the group-of-origin is ambivalent or hostile to the mobility of its members, necessitating a decisive abandonment of the past as a condition of mobility. Under these conditions, mobility upward will require an unusually high degree of personal motivation.

test a number of hypotheses based on these conceptions of the sources of mobility strivings. Specifically, the following predictions were made about upward- and downward-aspiring boys.[2]

1. *Energy level.* The upward-mobile boys possess a high degree of available energy for use in social and work activities. Downward-mobile boys, on the other hand, show a diminished vitality.

2. *Achievement mode.* The boys aspiring to higher status show a pervasive achievement orientation and a secure sense of their own effectiveness in reaching goals. The downward mobile are less oriented toward achievement and dominated by a concern with security.

3. *Orientation toward the future.* The time perspective of the upward mobile tends to be extended, while the downward mobile tend to have a constricted orientation in time.

4. *Personal standards.* The upward mobile boys manifest well internalized moral values and standards of personal behavior. Internalization is relatively incomplete among the downward mobile.

5. *Autonomy.* The upward aspiring display a precocious independence from the family, and a strong sense of autonomy in choosing values and goals. Downward mobile boys, on the other hand, are tied in a dependent relationship to the family.

6. *Orientation toward the self.* Upward-mobile boys are realistic in assessing themselves, and show a high degree of self-confidence. The downward aspiring reveal ambivalence toward the self and a lack of poise in social situations.

7. *Family milieu.* In this area, our predictions were firmer and more specific about the downward mobile group. Previous research (3) led us to expect a pattern of ambivalent dependency in the downward mobile boys' relations with their families. Rebellion and rejection of parental values combined with a strong dependency stem from inconsistent and punitive treatment at the parent's hand: in short, we anticipated an authoritarian family setting.

In addition to these specific predictions, we were interested in inquiring whether the relationship to the family, in upward mobile boys, is characterized by an ambivalent rejection of family values or by a conflict-free differentiation of the self from the family.

[2] In this paper we will use the terms "upward aspiring" and "upward mobile" interchangeably. It should be clear that we do not assume that upward aspiration will necessarily lead to successful social mobility.

METHOD

Subjects

The data for our analysis derive from a national sample survey of adolescent boys conducted by the Survey Research Center.[3] The sample for the total study consisted of a thousand boys in the 14- to 16-year age range selected by probability sampling methods. Each boy was given a personal interview at school by a member of the research center's field staff. Open-ended questions and projective questions were used, and interviews lasted from one to three hours.

In the present analysis, we have used as the base sample the interviews taken with urban nonfarm youth whose fathers' occupations fell in any of the following categories: small business owners, self-employed artisans, white collar, sales, and clerical workers; skilled and semi-skilled manual workers. The reason for excluding the sons of men occupying positions at the two extremes of the skill hierarchy becomes apparent when the mobility measure is discussed.

Mobility Aspiration

Each of our youthful respondents was asked the following questions:

a. What kind of work would you like to do as an adult?

b. Are you pretty sure about this or do you think you're just as likely to go into something else?

Each boy's aspiration was classed on the occupational scale and compared to his father's position to determine whether it was equivalent to or higher or lower than the father's job in the hierarchy of skills and status.

The distribution of aspiration types by father's occupation is presented in Table 1 for

[3] The survey was sponsored by the Boy Scouts of America. The complete questionnaire used in the study has been deposited with the American Documentation Institute. Order Document No. 5422 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for 35 mm. microfilm or $2.50 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
DISTRIBUTION OF ASPIRATION TYPES BY FATHER'S OCCUPATION

Aspiration	Father's Occupation			
	1 Professional, Managerial ($N = 136$) %	2 White Collar[a] ($N = 183$) %	3 Manual Skilled[b] ($N = 335$) %	4 Manual Unskilled ($N = 70$) %
Upward		46	58	84
Stable	67	29	34	16
Downward	33	25	08	

Note.—Ten per cent of the total urban sample of boys was lost in the process of rating the boy's aspiration in relation to father's occupation. These were *Ss* who had not decided what they wanted to be, who gave vague answers, or who wished to be farmers.

[a] This category includes small business owners in addition to sales and clerical workers.

[b] This category includes both skilled and semi-skilled workers.

all urban youth in the original sample. An obvious fact is illustrated in this table. Boys whose fathers currently occupy jobs in the top category are by definition limited to equivalent or lower status aspirations, just as those whose fathers are unskilled workers are barred from downward aspiration.

Because of this limitation on the relative freedom of direction in the aspirations of boys from families in Groups 1 and 4, and because we were interested ultimately in both upward and downward mobility, the sample was restricted to those boys from the two middle categories. The aspiration behavior of boys from the extreme groups may reflect in part a statistical artifact. That is, assuming that by chance a certain proportion of boys will aspire to a status-skill level different from the parent's, the direction of deviation is automatically determined for boys in Groups 1 and 4. In the middle groups, on the other hand, deviation is possible in either direction, and the selection of one or the other type should reflect certain specific motivational factors. One might maintain that the fact of aspiring to a level other than the father's in itself reflects unique psychological characteristics, but since the direction of deviation is a focus of our present concern, opportunity for choice in both directions was controlled.[4]

[4] The group used for analysis consisted of 335 boys from working-class background and 183 from white-collar families. Since the upward mobile group contained a disproportionate number of boys from blue-collar backgrounds, we ran all mobility analyses separately for the two broad occupational groups (Categories 2 and 3 in Table 1). Relationships were of approximately the same order within each of these groups. Analysis is presented for the combined group in order to avoid complicating tables unnecessarily, but in all cases relationships reported also held within each of the two background groups.

RESULTS

In the following pages we show contrasts among upward, downward, and stable groups. In most cases, we find linear relationships between status aspiration and other variables: the upward mobile are particularly high or low on a variable, the downward mobile are at the opposite extreme, and the stable group falls somewhere in between. Occasionally we find that either the upward or downward aspiring group clearly distinguishes itself from the other two groups. And in a few relationships, the stables stand out, the linear relationship gives way to a curvilinear one.

Since our predictions all concern the mobility categories and there was uncertainty about the exact position of the stable group, chi square was used to test the significance of variation among the three groups.

Energy Level

We see here a sharp contrast between upward and downward subjects (*Ss*): while the upward mobile boy is unusually lively and energetic, the downward aspirer is inactive and apathetic. On all of our measures, the upward group reveals a high commitment of energy to social and recreational pursuits, and yet not at the expense of work involvement; they also seem to show a spirited and enthusiastic love of activity. The downward mobile boy, in contrast, shows an impoverished vitality: he does not participate energetically or enthusiastically in recreational activities.

Looking first at boys' memberships in organizations (Table 2, Item 1), we find that three quarters of the upward mobile *Ss* belong to some organized group; while slightly less than half of the downward aspiring boys report membership. The upward mobile also have a relatively high proportion of multiple memberships, compared to the other two groups. A similar relationship appears in the area of general leisure activities: almost half the upward mobile report having tried twenty of more separate activities while only about one quarter of the downward mobile fall in

TABLE 2

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF LEVEL OF ENERGY

Item	Mobility: Upward (N = 277) %	Mobility: Stable (N = 168) %	Mobility: Downward (N = 73) %	χ^2*	p Level
1. Number of group memberships[b]					
a. none	25	33	52	26.91	.01
b. one	32	38	22		
c. two	23	20	17		
d. three or more	20	09	09		
2. Number of leisure activities					
a. fewer than 20	52	65	67	11.65	.01
b. 20 or more	48	35	33		
3. Dating					
a. do date	66	59	52	5.67	.10–.05
b. do not date	34	41	48		
4. Employment					
a. have jobs	49	47	51	.44	.90
b. do not have jobs	51	53	49		
5. Leisure reading					
a. do not read	17	27	25	7.93	.05
b. fiction-novels, mysteries	32	23	11	12.31	.01
c. travel and adventure	31	20	18	10.52	0.1
d. technical, scientific	09	04	01	11.13	.01
e. sports and hobby books	10	09	04	2.44	.30–.20
f. history, biography	08	07	05	.27	.90–.80
g. animal stories	05	04	05	.73	.70–.50
h. comics, joke books	21	26	27	1.95	.50–.3
i. newspapers, magazines	19	16	19	.96	.70–.50
6. Proportion of activities enjoyed					
a. fewer than half	45	41	59	6.09	.05
b. one half or more	55	59	41		
7. Additional activities desired					
a. suggest activity	63	31	10	87.67	.01
b. do not suggest activity	37	69	90		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[b] In this case, the chi square was derived from a 3 x 4 table with six degrees of freedom.

this category; the stable group is approximately midway between the other two (Table 2, Item 2). There is a small difference in dating, the upwards being slightly more likely, and the downwards least likely to date (Table 2, Item 3).

Although they more commonly engage in leisure pursuits, the aspiring boys are not significantly different from others in having job commitments. One might expect that the reason the downward-aspiring boys play less is that they work more. But this does not seem to be the case (Table 2, Item 4), the three groups being approximately equal in the percentage who hold jobs. The more active social life of the upward mobile boy represents an increment in total activity rather than a substitute for work activity. The upward mobile boy is more likely to report leisure reading. He has a more diversified reading pattern, and his reading includes more demanding material. The downward aspiring boy much less often reads novels and mysteries, travel and adventure stories, and technical, scientific literature. Stable boys fall between the other two groups on all specific categories of reading, but are like the downward mobile in the proportion of nonreaders. The single case in which the trend favors the downward mobile boy is in the category "comics, joke books" (Table 2, Item 5).

Upward mobile boys are not only more active currently but also seem to be more enthusiastic about their activities and more receptive to new experiences. After our *S*s had checked the leisure pursuits they had experienced within the last year, they were asked to check the ones they had especially enjoyed. In general, the two indices bear a low negative relation to each other—that is, boys who do more things are more selective in naming those they "particularly enjoy." Despite this tendency for a negative association, we find that the upward aspiring boys (who take part in more leisure activities) are nevertheless enthusiastic about the things they do. They are like the stables in the proportion of their leisure pursuits that they designate as particularly enjoyable; while the downward mobile, who have tried relatively few leisure activities, are much less enthusiastic about these few (Table 2, Item 6). When asked if there are other activities they would like to try, aside from those appearing on the list, almost two-thirds of the upward group name something else they would enjoy. This responsiveness contrasts sharply to the other groups: a third of the stable group, and only one tenth of the downward group suggest any

additional unlisted activities that appear to them (Table 2, Item 7).

Achievement Mode

We expected that the degree of striving reflected in a boy's job aspiration would be part of a general achievement orientation, that upward occupational aspiration would be associated with the pattern of behavior which McClelland *et al.* (7) have termed the "achievement syndrome," and that the upward mobile boy would show concern with achievement as against security and would be interested in meeting self-imposed standards of excellence. The downward aspiring boy, on the other hand, should show the same lack of vigor, the same constriction in setting other goals that we have seen in his choice of a future occupation.

The most direct evidence on these points is found in response to other questions in the occupation area. The *S*s were asked to choose among criteria for judging the attractiveness of a job, to choose the two most and the two least important dimensions. Compared to the stable and downward groups, upward mobile boys stress interesting work and have a less dominating concern about security. They are highest of all three groups in desire for status achievement. The downward mobile also strongly wish for status, but their status drive is not bound to a demand for interest in the content of the job; rather, they would most like to have status combined with security. The answers of our stable group are interesting on this item; relatively unattracted by status and fame, they stress security and interesting work (Table 3, Item 1).

When we ask boys to rationalize their own job choices, the upward aspiring *S* emphasizes interest in the work itself, while the downward mobile more often adduces the ease of the work, or the ease of obtaining employment in the field (Table 3, Item 2).

It would seem, then, that the upward mobile *S*s are more willing to yield security for the sake of achievement. In answer to a forced-choice question: "Which would you rather have—a job where you're sure you won't be laid off or one you can't be so sure of but where you have a chance to be a big success?" the striving boys choose the challenging, less secure job significantly more often than boys in either of the other groups (Table 3, Item 3).

Achievement orientation is often inferred from the willingness to forego immediate gratification for long-term goals. One indication of this appears in answer to the question, "What do you spend your own money for?" The boys who aspire high are more likely to say they save their resources for education or other long term goals (Table 3, Item 4). As we would expect, boys aspiring to higher occupations also reveal more extensive educational plans than do other boys (Table 3, Item 5).

TABLE 3

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ACHIEVEMENT MODE

Item	Mobility: Upward (*N* = 277) %	Stable (*N* = 168) %	Downward (*N* = 73) %	χ^2*	*p* Level
1. Job criteria selected[a]					
a. interesting work	56	51	33	15.74	.01
b. status achievement	44	28	36	22.15	.01
c. security	50	59	65	5.84	.10
2. Reasons for job choice[a]					
a. status	20	09	03	11.65	.01
b. ease of job or ease of obtaining job	04	12	32	15.27	.01
c. interest of work	70	78	58	5.99	.05
3. Preference: success or security					
a. success	52	35	33	12.00	.01
b. security	48	65	67		
4. Savings					
a. do save	30	21	18	6.16	.05
b. do not save	70	79	82		
5. Education plans					
a. beyond high school	69	42	12	140.06	.01
b. not beyond high school	31	58	88		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] Categories here are not exhaustive. Tests are run on the presence or absence of particular answers.

Orientation toward the Future

The upward mobile boys have an extended time perspective and a marked interest in attaining adult status; while the downward group shows constriction in time, as in other aspects of the life space. We asked our *S*s what decisions they would have to make in the next few years; the answers were rated for the breadth of time perspective. The upward

TABLE 4
RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ORIENTATION TOWARD THE FUTURE

Item	Mobility			χ^2	p Level
	Upward (N = 277) %	Stable (N = 68) %	Downward (N = 73) %		
1. Time perspective on decisions[b]					
a. within high school	17	18	29	5.21	.10
b. beyond high school	78	73	77	1.75	.50
c. distant future	53	51	26	16.51	.01
2. Sources of self-esteem[b]					
a. assuming adult role	31	23	16	6.91	.05
b. belonging, being part of a group	46	56	32	11.34	.01
c. being accepted by peers	03	07	14	10.90	.01
d. nothing	04	08	11	6.39	.05
3. Preference for coed or all-boy club[a]					
a. all boy	39	48	59	10.60	.05
b. neutral	23	16	14		
c. coed	38	36	27		

[a] Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In this case, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[b] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

aspiring boys are more likely to mention decisions in the more distant future, such as adult roles and goals. The downward mobile more often name immediate decisions, those within the high school period (Table 4, Item 1).

In answer to the question "What things that you do—at home, in school, or with your friends—make you feel important and useful?" the upward striving boy more often states instances in which he has assumed adult roles. He is less likely than the downward mobile *S* to mention peer acceptance as a source of self-esteem. An interesting finding here is that the stable group is highest of all three in choosing "belonging, being part of a group," one of the few occasions where this group does not fall somewhere between the other two. We suggest, though very tentatively, that this may reflect a concern with maintaining stable interpersonal ties.

We may also point to the apparently greater social maturity of the upward mobiles. As we have seen, they are somewhat more likely to date. They are also more eager for heterosexual social activities; when asked to choose between an all-boy or a coed social club, the upward aspiring less often prefer the former (Table 4, Item 3).

The Internalization of Personal Standards

Another point of difference between the groups concerns the interiorization of standards. Upward mobile boys consistently show a more complete internalization of personal controls. Downward mobile boys show a tendency to externalize standards and to rebel against them.

A set of projective pictures showed an adolescent boy in conflict between a promise made to his parents (to be home at a certain time) and peer pressure (to stay out later). Our *S*s were asked what the boy would do and how he would feel about it.

In all of our groups, approximately two thirds of the respondents say that the boy will go home. However, the motives attributed to the boy differ sharply. The upward mobile boys more often have the boy go home because "he promised" or because "his parents trust him," whereas the downwards say he will go because of a fear that the parents will find out and punish the transgression (Table 5, Item 1).

Regardless of their responses to these questions, all *S*s were asked: "If the boy decided to stay with his friends awhile, do you think he'd tell his parents about it later?" The upward aspiring group answer "yes" more frequently than other boys (Table 5, Item 2).

Another series of questions concerned boys' conceptions of and general reactions to rules. Here, again, the upward mobile group reveals a greater degree of internalization. When asked why a boy might break a rule, they focus on uncontrollable emergency situations and those in which a boy is old enough (i.e., responsible enough) to guide his own behavior. On the other hand, the downwards more frequently give responses fitting the categories of "rebellion against the parent," "irresistible impulse," and "when parental authority is not present" (Table 5, Item 3). When asked what kind of a rule they would never break, upward mobile *S*s more often mention those which involve responsibility to others; they less frequently say that there is no rule, or no rule that comes

TABLE 5

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF INTERNALIZATION OF PERSONAL STANDARDS

Item	Mobility: Upward (N = 277) %	Stable (N = 168) %	Downward (N = 73) %	χ^2[a]	p Level
1. Reaction to parent-peer conflict[b]					
a. adheres to promise because of sense of trust	37	19	16	24.03	.01
b. Adheres to promise because of fear of punishment	04	06	14	6.38	.05
2. Honesty with parents					
a. would tell parents	55	42	44	9.15	.05
3. Conditions for breaking a rule[b]					
a. emergency	18	10	10	6.56	.05
b. boy mature enough	11	07	01	6.49	.05
c. rebellion	10	14	22	10.11	.01
d. impulse	05	11	16	11.41	.01
e. parental authority absent	05	10	27	31.50	.01
4. Unbreakable rule					
a. no unbreakable rule	03	17	21	10.85	.01

[a] Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[b] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

to mind, which they wouldn't break (Table 5, Item 4).

Autonomy

We saw earlier that the striving boys are interested in assuming adult-like roles, those, that is, in which they themselves are responsible for their behavior and for a job. We see here a seemingly greater drive towards independence and responsibility. In contrast, the downward mobile, though they show signs of rebellion against the parents, are also more dependent on them.

In one series of items, our *S*s were asked whose advice they would prefer to take on particular problems—their parents' or friends'. The issues ran from such central ones as what time to be in at night to matters of taste, such as personal grooming. While there were no very striking differences in the proportions who would heed their parents on any of these topics, one interesting difference appears in the proportion of boys who inject a note of independence while discussing these issues. Thirty per cent of the striving group spontaneously assert that they would follow their own ideas on at least one of the six issues, and only three per cent of the downwards give such a response (Table 6, Item 2).

We find another indication of independence from the family in response to the question, "Can a friend ever be as close as a family member?" The upward mobile *S*s most often agree to this idea (Table 6, Item 3).

When our *S*s are asked what the role of a club leader ought to be, we again find that the aspiring boys desire independent direction of their behavior. They limit the leader's function to that of a helper (Table 6, Item 4). A coder rating based on this question and on one which asked the *S* to describe an adult leader he had liked, revealed that the downward mobile *S* is

TABLE 6

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF AUTONOMY

Item	Mobility: Upward (N = 277) %	Stable (N = 168) %	Downward (N = 73) %	χ^2[a]	p Level
1. Independence in allocating funds					
a. no independence	01	01	19	59.27	.01
2. Advice on decisions					
a. interjects own opinions	30	21	03	24.69	.01
3. Intimacy of friendship					
a. can be as close as family relationship	59	54	34	9.92	.01
4. Role of adult leader[b]					
a. helper	58	51	38	9.71	.01
b. decision maker	40	43	51	3.67	.20
5. Authority reliance					
a. high	38	47	54	6.06	.05
b. moderate	62	53	47		
6. Adult ideal[b]					
a. family member	41	50	55	5.99	.05
b. unrelated adult acquaintance and composite	31	24	18	6.74	.05

[a] Unless noted, the chi square was derived from a 3 x 2 table with 2 degrees of freedom.

[b] Categories here are not exhaustive of responses given to open questions, and more than one response was often given. Tests are run on the presence or absence of particular answers.

more authority reliant than other boys (Table 6, Item 5).

Finally, a crucial index of independence from the family is yielded when *S*s are asked to choose an adult ideal. As we would expect, the stable group, which includes all boys who choose the same occupation as the father's, also has the largest proportion of boys who choose the father as an ideal. The upw rds more often choose a model outside the f mily, or describe a composite of characterist com several people. More often than other boys, downward mobiles say there is no adult they wish to be like. And despite the fact that they choose the father less often than stable boys, they give the largest proportion of within-family models of any group, particularly choosing grandfathers and uncles. Here again we can observe the ambivalence of the downward mobile boys; though they are covertly rebellious toward family authority, they cling to within-family models. Another way of viewing this finding is to see in it a reflection of the downward mobile boys' narrowed and immature life space. Havighurst (4) has found that the tendency to choose within-family models characterizes the earlier years of adolescence. Our downward mobile boys retain the more infantile, more restricted image of the adult world.

The upward mobile boys, on the other hand, tend to a certain precocity in the loosening of family ties. Yet we have seen in the previous section on internalization that this is not associated with rebellion from or rejection of the family. Further evidence along these lines appears in the discussion of family milieu.

Until now evidence has been presented bearing on the *sense* of autonomy. Some other data suggest that these differences are related to the objective degree of autonomy permitted upward and downward boys by their parents. We asked: "What are your parents' ideas about the way you spend your money?" One fifth of the downward mobile respondents replied that they do not have funds of their own but ask their parents for money as they need or want some special thing. In the upward mobile group, fewer than one per cent gave this answer (Table 6, Item 1). Since there is no difference between the groups with respect to actual spending of money, we may infer that the difference resides in the *control* of spending.

Orientation Toward the Self

Upward mobile boys show a high degree of self-acceptance, and a confidence in social situations. The downward mobile is more ambivalent toward himself and more unsure and conflicted in social interaction.

Interviewers rated boys on a number of variables, among which were self-confidence, humour, and the clarity with which they organized and presented their opinions and attitudes. We may interpret these ratings, with caution, as reflecting the boy's social skill and ability to handle himself in social interaction with an adult. On each of the ratings, the upward mobile are high; they are more self-assured, show humour more often, and are better organized in the interview. The downward mobile are relatively unconfident, humorless, and disorganized (Table 7, Items 1, 2, and 3).

We find signs of self-rejection and demorali-

TABLE 7

RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF ORIENTATION TOWARD THE SELF*

Item	Mobility			χ^2*	p Level
	Upward (N = 277) %	Stable (N = 168) %	Downward (N = 73) %		
1. Self-confidence*					
a. high	30	23	18	7.93	.10
b. average	52	57	57		
c. low	18	20	25		
2. Humour					
a. present	74	67	58	7.14	.05
b. absent	26	33	42		
3. Organization of ideas*					
a. high	60	53	50	7.31	.10
b. average	22	27	20		
c. low	18	20	30		
4. Desired changes[b]					
a. self rejecting, major	06	10	15	6.36	.05
b. changes that are impossible	06	07	23	11.33	.01
c. changes within boy's own power	37	29	05	14.10	.01

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

* In these cases, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[b] Categories here are not exhaustive of responses given to an open question, and more than one response was often given. Tests are run on the presence or absence of particular answers.

zation in the downward mobile boys' answers to the question, "What would you like to change about yourself if you could—about your looks or your life or your personality?" They more often desire changes so gross or so central as to indicate alienation from the self; and they more often wish for changes that are unlikely to occur. The upward mobile boy more often refers to changes he has the power to effect himself. He is more realistically critical of himself, and less self-rejecting (Table 7, Item 4).

Family Milieu

The reader may recall that we felt far more confident of our predictions about the downward than the upward mobile in this area. We believed that the downward group would show a pattern of ambivalence towards the parent, a mingling of dependency and hostility; and we further held that this pattern would arise out of inconsistent, overly harsh methods of discipline. We did not, however, feel we could make any specific statement about the upward mobile: here we were interested in discovering what we could about the family milieu, in particular whether upward aspiration is associated with a rejection of family values, or a defensive identification with a forbidding parental authority, or by that relatively amiable relationship to the parent which permits the growth of autonomy.

TABLE 8
RELATIONSHIPS BETWEEN MOBILITY ASPIRATION AND INDICES OF FAMILY MILIEU

Item	Mobility			χ^2*	p Level
	Upward (N = 277) %	Stable (N = 168) %	Downward (N = 73) %		
1. Punishment[a]					
a. physical	02	08	15	19.12	.01
b. deprivational	66	69	65		
c. psychological	32	23	20		
2. Portrayal of parental figures					
a. harsh	27	35	48	11.59	.01
b. non-harsh	73	65	52		
3. Disagreements with parents					
a. have disagreements	66	56	49	6.39	.05
b. do not have disagreements	34	44	51		
4. Parents' attitudes toward way boy spends money					
a. disapproval	19	16	05	8.58	.05
b. no disapproval	81	84	95		
5. Leisure activities with parents[a]					
a. share many	19	11	09	17.23	.01
b. share some	73	74	68		
c. share none	08	15	23		

* Unless noted, the chi square was derived from a 3 x 2 table with two degrees of freedom.

[a] In these cases the chi square was derived from a 3 x 3 table with four degrees of freedom.

To begin with, we find that deprivation seems to be the dominant method of discipline for all groups; differences here are not significant. Striking differences appear, however, in the relative frequency of "psychological" and "corporal" methods of punishment. Only one in fifty of the upwards say that their parents use physical punishment; approximately one in ten of the stables and one in seven of the downwards report this method. In contrast, the use of "psychological" techniques (such as "given a good talking to") occurs in more than one quarter of the upward mobile responses, about one in five of the stables, and only one in eighteen of the downwards (Table 8, Item 1). We find that the parents of the upward aspirers avoid the harshest method of punishment; the "psychological" methods, which we suspect encourage the child's internal controls, are infrequently used by the parents of downward mobile boys.

Are our mobile *S*s presenting an overidealized picture of their families? We looked for the pattern of repressed hostility and surface idealization which has been shown to characterize defensive identifiers (2, 5, 9). But our data reveal that the opposite is true: in the projective materials—where the *S*s are not discussing their own parents and where they have an opportunity to discharge repressed hostility without danger—the upward aspiring portray parents as less harsh than do the other groups (Table 8, Item 2). And as we shall see, they are more aware of differences between themselves and their parents. They more commonly report that their parents have some "old-fashioned ideas, or ideas they disagree with," (Table 8, Item 5), and that their parents in some respect disapprove of the way in which they spend their own money (Table 8, Item 6). We saw previously (Table 6, Item 3) that they are not likely to feel that family relationships are necessarily more intimate

than friendships; and they less often choose an ideal adult from within the family group (Table 6, Item 6). The upward mobile boys are, then, more sensible of differences with parents.

Do the aspiring boys have a detached rather than congenial relationship to their parents? The single bit of evidence on this question suggests that this is not the case: they more often report engaging in leisure activities with their parents (Table 8, Item 5).

The downward mobile boy's family relationship is, as we expected, marked by ambivalence: the pattern is one of conscious idealization accompanied by unconscious suspicion and hostility.

We have already seen that the parents of this group are more often severe in their punishment. And in this group we find evidence of repressed hostility: in the projective stories, the downward mobile boys picture parental figures as both harsh and suspicious (Items 3 and 4, Table 8). Yet consciously they are more likely to deny differences with their parents. Some of our findings on autonomy have shown this; and in addition we find that these boys most often say they have no disagreements with their parents (Table 8, Item 5), and they least often say their parents disapprove of their handling of funds (Table 8, Item 6). The downward mobile boy shares leisure activities with his parents somewhat less than do boys in the other two groups (Table 8, Item 7).

Alternative Hypotheses

Before we discuss the possible meanings of these findings, we must consider briefly alternative hypotheses which might account for them. These are: (*a*) An age differential among the three groups; thus, the upward aspiring *S*s might have given "little boy" job hopes, while the non-mobile group presented more mature aspirations; (*b*) a difference among the groups in frame of reference, so that the upward aspirers may have understood the interviewer to be asking for "fantasy" aspirations while other boys took the question to mean "realistic expectations"; (*c*) objective differences in the social status of the groups: perhaps the upward mobile boys' families are at the upper end of their group in income or education, and so provide their sons with an objectively greater opportunity for mobility; (*d*) a difference in intelligence among the groups.

First, on the question of realism. There were in our total sample fewer than one per cent of the boys who gave us their job choices distinctly "little boy" or "glamour" occupations. Among boys in the mobility analysis subsample, there were no aspirations listed which were obviously or totally based on ideas of glamour. When the boys are asked how sure they feel about the job they've chosen, we find that very few of them say they're not at all sure; more important, the variation that does occur is not related to the type of aspiration they express. Boys in the three groups show approximately equal certainty that they will *get* the jobs for which they are aiming (Table 9, Item 1).

Furthermore, we find a difference in educational plans, one which is appropriate for the difference in job aspirations; this indicates at least a minimal degree of realism in our *S*s'

TABLE 9

RELATIONSHIP OF MOBILITY ASPIRATION TO AGE, CERTAINTY ABOUT ASPIRATION, FAMILY INCOME, FATHER'S EDUCATION AND INTELLIGENCE

Item	Mobility			χ^2*	p Level
	Upward (N = 277) %	Stable (N = 168) %	Downward (N = 73) %		
1. Certainty about aspiration					
a. high	57	58	54	3.22	.70
b. moderate	34	33	33		
c. low	09	09	13		
2. Age					
a. fourteen	33	29	31	1.94	.80
b. fifteen	30	36	33		
c. sixteen	37	35	36		
3. Family economic status[a]					
a. high average	32	35	31	2.76	.70
b. average	53	49	52		
c. low	15	16	17		
5. Father's education					
a. grade school	32	27	25	3.07	.70
b. high school	50	58	55		
c. college	18	15	20		
5. Verbal ability					
a. high	23	11	03	33.92	.01
b. average	69	75	74		
c. low	06	14	23		

* Unless noted, the chi square was derived from a 3 x 3 table with four degrees of freedom.

[a] The measure used for economic status was the Remmers House and Home Scale.

understanding of work preparation (Table 3, Item 5).

The groups differ neither in age (Table 9, Item 2) nor in present socioeconomic status. The upward mobile boys are not from homes which are higher economically or in educational background (Table 9, Items 3, 4).

The two rough measures of intelligence available both show the striving group to be somewhat superior. The upward aspiring boy is more likely to report leisure reading (Table 2, Item 5), and he is rated by coders as more facile verbally (Table 9, Item 5). The downward mobile are rated low on verbal skills, but their reading pattern does not differ significantly from that of the stable group ($\chi^2 = .10$; $p > .90$ with 1 *df*). We will defer a consideration of this until the discussion.

DISCUSSION

We may interpret our data on the upward aspiring boy in relation to ego structure and functioning. We did not, of course, make a direct investigation of intrapsychic processes; furthermore, our understanding of "normal" ego functioning is, at present, limited and tentative. Nevertheless, it may be of some interest to attempt a formulation of our results in the current vocabulary of ego psychology.

Among upward aspiring boys we find a high level of diversified activity, suggesting that energy is available to the ego for focused use in work and play. We may infer that the antecedent condition for this accessibility is in the ego's having at its disposal a relatively high degree of neutralized energy. If we assume that both groups, mobile and non-mobile, do not differ in primary energy resources, we may conjecture that the aspiring *S*s have less energy tied to the resolution of conflict, and consequently gain in neutralized, disposable energy.[8]

Another sign of the comparative effectiveness of ego functioning among aspiring boys is seen in the greater degree of cohesion among the three major systems within the personality. On the one hand, ego processes in the mobile boy are less likely to be imperiled by the breakthrough of impulses (When would a boy break a rule?); on the other hand, he is less likely to show evidence of an excessively punitive super-ego. The greater degree of internalization of moral values among aspiring boys suggests a more harmonious articulation between the ego and super-ego systems. Our data show that this integration is not purchased at the expense of an unusual quantity of guilt.

A consideration of the intelligence factor is relevant here. Although we had no direct measure of intelligence, we saw that the upward mobile boys read more and read more difficult things; and they are judged on the basis of the interview to be more verbally skilled. In all likelihood there *is* an intellectual difference among our groups; one plausible interpretation of many of our findings is that they merely reflect group differences in IQ. We feel, however, that such a formulation assumes an overly simple causal sequence—for one thing it fails to clarify the mechanisms which mediate between IQ and other indices of successful ego functioning. A good deal of recent data points to a high degree of association between intelligence and variables which indicate effective ego functioning. The direction of causality and connection among these variables is by no means understood. Furthermore, to explain our data as a function of intelligence alone would not allow us to generate predictions in the area of family relations. If we assume, however, that our data are best understood by positing differences in intrapsychic organization—and particularly in ego structure and functioning—we may then posit specific expectations about family milieu.

We would expect—on the basis of the higher energy level and the more refined control of impulsivity among upward mobile *S*s—that they have relatively good relationships within the family and have successfully resolved infantile object ties. We mean by this that there is a relative absence both of covert hostility toward the parents, and of overt dependence upon them, together with a relatively high degree of objectivity about family relations.

We have seen that this is, in fact, the case. The mobile boy asserts (and probably has been encouraged toward) greater independence of judgment and behavior. It will be recalled that

[8] Cf. Rapaport (8, p. 353): "In other words, the amount of energy which the person can dispose by investing it in objects, by becoming interested in activities, even when essential drive-aims and drive-objects are in abeyance, is an indicator of ego-autonomy and ego strength."

he expresses a desire for equalitarian rather than dependent relationships with adults. He frequently uses his own convictions as authority in making decisions.

The aspiring boy more often admits differences between himself and his family. He seems able to assert the legitimacy of an autonomous self-definition—to the extent, at least, of selecting an ego ideal outside the family setting.

The independence we have observed has not occurred through rebellion from or rejection of the family. Despite their ability to differentiate themselves from their parents, the upward aspirers give evidence of a congenial relationship with them. They share their leisure activities; in response to a projective measure they reveal relatively little covert hostility toward parental figures.

Their parents are, in fact, comparatively lenient; they are more likely to employ mild and essentially verbal discipline and use physical punishment infrequently.

We have seen, then, that the upward aspiring boy is characterized by a high energy level, the presence of autonomy, and a relatively advanced social maturity. These attributes may be viewed as derivatives of a generally effective ego organization, one which has developed out of the successful resolution of infantile object-ties and conflicts. We would expect, on theoretical grounds, that this development would be most likely to occur in a family setting where the parents allow the child a nonambivalent connection with them, where autonomy can emerge without conflict. As we have seen, the data we have suggest that this is the case.

We have already presented, in the introductory section, our view of the internal psychic processes which characterize downward mobile boys. From the specific perspective of ego functioning, we see an apparent blocking or impoverishment of energy which should, ideally, be available to the ego. We may infer that neutralization is not being successfully accomplished. There is a relatively poor articulation among the psychic systems: impulses threaten the ego's integrity; the superego seems overly severe and yet incompletely incorporated. These boys seem humourless, gauche, disorganized—relatively so, at least. Perhaps the most telling and poignant datum which the study locates is their response to the possibility of personal change, their tendency to want to change intractable aspects of the self, and the degree of alienation revealed by their desire to modify major and fundamental personal qualities. As we have seen, this pattern is associated with a certain quality of family interaction: the boy gives evidence of an ambivalent tie to the parents, a mixture of overt dependency and covert aggression. One of the determinants of this ambivalence, we may infer on theoretical grounds, is in the parents' relatively punitive style of discipline which fails to establish self-governing controls in the child.

A final point concerning the mechanisms through which the differing aspirations become established: In the case of upward aspiration we assume a direct connection between the socialization process and the aspiration to social mobility. That is, we assume that the parents transmit the value on mobility during socialization, and that the child incorporates this as part of a complex of values which includes autonomy and general achievement. In contrast, we conjecture that the process in downward aspiration may be more indirect: the rebellion engendered by an ambivalent family atmosphere may be directly expressed in a rejection of general social goals; or we may have a diffuse demoralization which causes (and prepares the child for) failure and isolation in social situations. In turn, these failures result in a general narrowing of goals, and a retreat from success; a central expression of this is the lowering of occupational goals.

SUMMARY

We have only recently begun to recognize that prevalent views of the dynamics of mobility, based as they are largely on literary sources and on extreme instances, require refinement and differentiation. The paper begins by distinguishing varieties of mobility behavior. There is presented a model of the psychodynamics of upward and downward aspiration, which is tested with data from a national sample of adolescent boys. The general formulation holds that upward mobility is found among boys with effective, autonomous ego functioning; downward mobility is seen as a symptom of demoralization. Psychoanalytic ego theory provides the framework

from which we make specific predictions of differences in the areas of activity level, achievement mode, time perspective, internalization of values, autonomy, self-esteem, and family milieu. The findings reveal sharp differences, in the predicted directions, between upward and downward aspiring groups.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, MARIE. *Anti-Semitism and emotional disorder*. New York: Harper, 1950.
2. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D., & SANFORD, N. *The authoritarian personality*. New York: Harper, 1949.
3 GOUGH, H. Personal communication, 1955.
4 HAVIGHURST, R. J., ROBINSON, MYRA Z., & DORR, MILDRED. The development of the ideal self in childhood and adolescence. In J. Seidman (Ed.), *Readings in adolescence*. New York: Dryden Press, 1953.
5. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
6. LYNN, K. S. *The dream of success*. Boston: Little, Brown, 1955.
7. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
8. RAPAPORT, D. *Organization and pathology of thought*. New York: Columbia Univer. Press, 1953.
9. SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-Semitism among Jews. *J. Pers.*, 1952, **20**, 199–218.

Received March 4, 1957.

THE RELATION OF ACHIEVEMENT MOTIVATION TO PROBLEM-SOLVING EFFECTIVENESS[1]

ELIZABETH G. FRENCH AND FRANCIS H. THOMAS

AF Personnel and Training Research Center, Lackland Air Force Base

THIS study was designed to test the hypothesis that achievement motivation is a factor in problem-solving effectiveness. If motivation is assumed to have both energizing and directing powers, then an individual with high achievement motivation should both work harder and be more oriented toward identifying and utilizing cues relevant to the obtaining of an achievement-related goal than one without such motivation. A previous study by Thomas (2) demonstrated that achievement motivation was related to the length of time a subject would work on a problem without objective knowledge of progress but it did not show that it was related to reaching a successful solution. It was the purpose of the present study to test the hypothesis that achievement motivation is related to obtaining a problem solution as well as to the length of time spent on the problem.

METHOD

Measure of Motivation

The measure of achievement motivation was a projective device called The Test of Insight (1) in which the *S* is asked to provide "explanations" for briefly described behaviors presented as typical of hypothetical individuals. The number of achievement-relevant explanations make up the score.

The Problem

The problem-solving apparatus consisted of a panel containing seven lights arranged in a horizontal row with a switch below each light. The wiring of the panel was such that some switches operated one light; some, two; and some, three. Similarly, some lights were controlled by one switch only and some by two or three. Operation of a switch always changed all the lights to which it was connected. Thus, if Light 1, controlled by Switch 1, was on and Light 3, also controlled by Switch 1, was off, operation of Switch 1 would reverse the pattern. *S*'s problem was to create a specified pattern of lights.

Subjects

The *S*s were 100 airmen undergoing basic training at Lackland Air Force Base. All *S*s selected had Armed Forces Qualification Test (AFQT) scores in the top 11 per cent of the airman population at large, in order to minimize the effect of intelligence on problem solution.

Procedure

Each day's *S*s, about 10 in number, were first seen as a group and given an orientation to the problem-solving that clearly indicated that good performance could be considered relevant to achievement goals. Then they took the test of Insight, described to them as unrelated to the main part of the experiment.

Upon completion of the test of Insight, each *S* was taken into the problem-solving situation individually. *E* demonstrated the apparatus to *S* and explained the important points, i.e., that switches might operate more than one light and vice versa, and that operation of a switch always changed all the lights it controlled. Each *S* was then given a practice period in which he was instructed to push the switches in order, one at a time, until he could demonstrate by reciting the combinations to *E* that he knew which switches controlled which lights. The time required for this varied from two to 12 minutes with a mean of 5.72; and the number of trials, from two to 11 with a mean of 4.98.

When *S* had demonstrated that he was in possession of all the necessary information to solve the problem, he was given the following instructions:

Now I am going to give you the problem which will be to indicate which switches must be in the "up" position in order to produce a particular pattern of lights. But first I want to explain how you must do this. In working out the solution you may operate the switches if you like but you may have only one of them in the up position at a time. In other words, you may not work out the complete solution on the board. You must figure it out in your head from what you know of each switch and its lights. When you think you have the answer, write the numbers of the correct switches on a sheet from this tablet and drop it in the box. Write only the answer; you may not figure on the paper. You will have 35 minutes in which to solve the problem. You may use as much of that time as you like and you may submit as many answers as you like in that time. Put each answer on a separate sheet and drop each in the box before continuing. Please number the sheets in order.

Is the procedure clear? You are to list the numbers of the switches which must be in the 'up' position to give the pattern I will ask for; you may submit as many answers as you like in the 35-minute time limit; and if you operate the switches in working out your solution you may have only one in the 'up' position at a time. If there are any questions ask them now because I

[1] This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the Research and Development Program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

can't answer them after I have given you the problem.

The problem is, "What switches must be in the 'up' position to have the greatest possible number of lights lighted?"

E then started a timer and *S* began to work. *E* recorded the time and number of each solution as it was dropped into the box. *E* did not look at the solutions given and *S* had no external indication as to whether he was right or wrong. Those *S*s who had not stopped voluntarily by 34 minutes were given a warning that the time would be up in one minute in order to permit them to write down a solution if they cared to.

The actual solution required the operation of five of the switches, which lighted all lights. The solution could not be reached with either of the other two switches in the "up" position. The solution also involved turning off lights already on and getting them on again with another switch. The switches could be operated in any order.

In summary, *S*s in possession of all the information necessary to solve the problem and of approximately equal intellectual ability but varying in level of achievement motivation were given up to 35 minutes to solve a problem presented to them as a measure of competence.

RESULTS

The analysis was based on the data from 92 *S*s. Two were discarded because their records indicated they had failed to follow instructions. The records of six others suggested that they had given a correct solution by chance; i.e., a correct solution was given in the middle of a long series of incorrect ones and was not repeated. Since these *S*s could be classed as neither solvers nor nonsolvers, they also were eliminated. An *S* was considered a solver if he submitted one correct solution and stopped work or if he continued to work after his first solution, submitting repetitions or alternate versions. Since the five necessary switches could be operated in any order, any one of the possible orders was correct. Some *S*s gave several of them. Of the 92 *S*s remaining, 39 solved the problem and 53 failed to solve it. Reaching the solution was unrelated to practice time (r_{bis} .05), number of practice trials (r_{bis} .05), AFQT (r_{bis} .18), time spent on the problem (r_{bis} .13), using all the time (chi square .01) or number of solutions given (r_{bis} .02).

TABLE 1

NUMBER SOLVING IN HIGH AND LOW MOTIVATION GROUPS AND MEAN MOTIVATION SCORES FOR SOLVERS AND NONSOLVERS

Attainment of Solution	Achievement Motivation		Motivation Scores	
	High	Low	Mean	*SD*
Solvers	25	14	9.28	3.95
Nonsolvers	22	31	7.34	3.62
	$\chi^2 = 4.45$		$t = 2.31$	
	$p = .05$		$p = .03$	

TABLE 2

MEAN TIME WORKED AND NUMBER STOPPING BEFORE THE TIME LIMIT FOR HIGH AND LOW MOTIVATION GROUPS

Achievement Motivation	Time Worked		Stopped before Limit	Worked until Limit
	Mean	SD		
High	27.14	10.02	25	22
Low	13.79	8.13	44	1
	$t = 6.99$		$\chi^2 = 23.22$	
	$p = .001$		$p = .001$	

To test the main hypothesis, the *S*s were divided into groups with high and low achievement motivation by dichotomizing the distribution of motivation scores at the median. The 47 *S*s with scores of eight or above were called the high motivation group; the 45 with seven or below, the low motivation group. The proportion of solvers and nonsolvers in each motivation group is given in Table 1. Although approximately the same number of *S*s solved and failed in the high motivation group, less than a third of the low motivation group solved. The chi square for this analysis is significant beyond the .05 level. Table 1 also gives the achievement motivation means for the solvers and nonsolvers. The difference between them is significant at the .03 level. The relation between achievement motivation and attainment or nonattainment of a solution may also be expressed by a biserial correlation, which proved to be .30. These data were considered to support the hypothesis that individuals with high achievement motivation are more likely to succeed in solving a problem that is related to achievement goals than are individuals with low achievement motivation.

Since the *S*s were permitted but not required to work for 35 minutes, and the actual time worked ranged from three to 35 minutes, it was possible to check the earlier finding that motivation was related to length of time the *S*s worked. The data for the relation between motivation and time worked are presented in Table 2. The mean time worked for the high motivation group was 27.14 minutes as compared to 13.79 minutes for the low motivation group. The correlation between motivation

score and time worked was .41. When time worked is considered in terms of whether the *S* stopped before the time limit or continued until stopped by *E*, the results are even more striking. Only one of the 45 low motivation *S*s worked until stopped, as compared with 22 of the high motivation *S*s, yielding a chi-square value of 23.22. The biserial correlation between motivation score and stopping versus not stopping was .54. Data for number of solutions offered by high and low motivation groups were essentially the same.

This relation between achievement motivation and time worked does not mean that the relation between motivation and problem success is solely a function of the fact that the more highly motivated *S*s worked long enough to arrive at a solution. In the first place, among those who did reach a solution, time spent on the problem was not necessarily the same as the time to the problem solution. Some *S*s continued to work beyond the first solution as described earlier. Thus, the mean time worked for the high motivation solvers was 25.2 minutes as compared to 18.7 for the low motivation solvers, a difference significant at the .05 level. However, the mean *solution times* for the two groups were 19.7 and 18.4 minutes, respectively, a difference that is not significant. In approximately the same period of time, 25 high motivation *S*s as opposed to 14 low reached the solution. Also, as pointed out earlier, there was no overall relation within the time limits of the experiment between the length of time worked and successful solution.

Nevertheless, the fact that highly motivated *S*s work longer obviously does have general implications for prediction of success. Consider, for example, the performance of the nonsolvers in the two motivation groups. All of the 31 nonsolvers in the low motivation group stopped before the time limit was up. Since they had stopped, there was no chance that they would reach a solution. Of the 22 nonsolvers in the high motivation group, 13 were still working at 35 minutes. Since the distribution of solution times ranged up to 35 minutes, it is reasonable to suppose that some of these nonsolvers would have reached a solution if permitted to work longer.

Three other differences between the high and low motivation groups were quite clearly suggested although they were not specifically predicted. The first is relevant to the point just discussed. Although there was no relation between time worked and reaching a solution ($r_{bis} = .13$), for the group as a whole the biserial correlation for the high motivation group only was $-.28$ and that for the low, .53 (Table 3). The former is significant at only about the .10 level, but the latter is significant at beyond the .01 level, and the difference between them is large enough to be suggestive of a real difference. The majority of the high motivation group but only a few of the low motivation group used most or all of the time, and in each group approximately equal numbers of the long working *S*s solved and failed to solve the problem. The big difference between the high and low motivation groups occurred among the *S*s who used only part of the time. Of those high motivation *S*s who worked a relatively short time, nearly all solved the problem. Only a small portion of the low motivation *S*s, who worked for a short time, solved.

TABLE 3
BISERIAL CORRELATIONS BETWEEN SOLUTION AND TIME, AFQT AND MOTIVATION FOR THE MOTIVATION GROUPS SEPARATELY

Solution vs:	High Achievement Motivation	Low Achievement Motivation
Time	−.28	.53
AFQT	.36	−.08
Motivation	.04	.41

A second difference between the groups not specifically predicted was the relation between AFQT and problem success. The AFQT range had been kept small so that performance would not be affected by differences in intelligence and as stated, the relationship is represented by a nonsignificant biserial correlation of .18. For the high motivation group only, even with the restricted AFQT range, the correlation was .36 ($p = .05$). The correlation for the low group was $-.08$ (Table 3). The difference here is again large enough to indicate a point for further investigation. The suggested interpretation is that the *S*s in the highly motivated group are all trying hard so that even small differences in ability become important. On the other hand, in the low motivation group ability differences are not important because the *S*s are not using their potential as effectively as they could. This finding needs to be checked over a wider

ability range before much can be said with assurance.

An additional finding which corroborates the notion that the high motivation group is more or less uniformly motivated, regardless of difference in scores, is a correlation of .04 between motivation score and solution for the high group. The low group had a correlation of .41 between motivation and solution (Table 3). About half the *S*s at each interval above the median solved the problem, while the proportion solving decreased with each interval below the median. Apparently, for this particular task, which had some intrinsic interest for most *S*s, achievement motivation of only about median intensity was necessary for optimum performance. A less interesting task might require a higher level or a more interesting one, a lower.

These last findings can be considered merely tentative, but they do point the way to fruitful areas for further investigation.

SUMMARY AND CONCLUSIONS

The hypothesis that when a problem is related to achievement goals, achievement motivation is related to problem-solving effectiveness and to the length of time *S*s will work on a problem without knowledge of results was tested. *S*s with high and low achievement motivation scores worked on a problem after first learning all the separate facts necessary for solution. They were permitted up to 35 minutes.

The results showed that *S*s with high achievement motivation were more likely to reach a solution than those with low motivation. High motivation *S*s also worked longer than low motivation *S*s. The relation between motivation and success was shown not to be a result of the relation of motivation and time worked since actual solution time for those who solved was not different for the two motivation groups. Other findings not specifically predicted are discussed and tentative interpretations are offered.

REFERENCES

1. FRENCH, ELIZABETH G. Development of a measure of complex motivation. *Res. Rep. 56–48*, AF Personnel and Train. Res. Center, April 1956, Lackland Air Force Base, Tex.
2. THOMAS, F. H. Visualization, experience, and motivation as related to feedback in problem solving. *Amer. Psychologist*, 1956, **11**, 444. (Abstract)

Received March 25, 1957.

THE EFFECT OF BRIEF SOCIAL DEPRIVATION ON BEHAVIORS FOR A SOCIAL REINFORCER

JACOB L. GEWIRTZ[1]
National Institute of Mental Health
AND DONALD M. BAER[2]
University of Washington

DOES the incidence of behavior for social contact with other people increase following a period of social deprivation? Stated in more general terms, are there social drives that function in a manner similar to the primary appetitive drives? This is the central question to which this investigation addresses itself, employing young children as subjects (*S*s).

Short-term procedures (other than learning operations) that account for variance in the effectiveness of a reinforcer (or in behavior output) are ordinarily labeled *drive* operations. One notable operation of this type in the case of the primary appetitive drives evolves from the fact that relevant stimuli (e.g., food, water) acquire maximal reinforcing value for an organism only subsequent to its recent deprivation of them. Other reinforcers of no apparent biological importance, like that provided by the opportunity to make a brief observation response, appear also to be raised in effectiveness following preceding periods of deprivation (2). In this context, it is a provocative question whether *social* reinforcers (those dispensed by people), postulated to possess reinforcing value through a history of conditioning, respond in a similar manner to deprivation. The relevance of this hypothesis accrues both from these precedents and also from the extensive use of the deprivation concept in a number of conceptual systems dealing with general behavior (e.g., 11, 13, 18, 19) and social behavior (e.g., 1, 7, 17, 20).

There is a considerable body of evidence (3, 8, 12, 21, 22) showing that an experimenter can increase the relative frequency of an arbitrarily chosen response of a child by dispensing certain social stimuli (words and phrases) contingent upon that response. In particular, the social reinforcer[3] *approval* is readily dispensed through the use of simple words and phrases like "Good!" "Fine!" and "Mm-hmm." Furthermore, approval appears to be representative of the reinforcers that control the purely social initiations made by children to adults.

For these reasons, this study is concerned with the possible responsiveness of the social reinforcer *approval* to a condition of social isolation, representing the deprivation of all social reinforcers (including approval).

METHOD

Each *S* served twice, once under a condition of Deprivation and once under a condition of Nondeprivation. Four discontinuous independent variables representing levels of Deprivation, the Sequence of the Deprivation and Nondeprivation conditions, the Sex of *E*, and the Sex of *S*, generated a 2 x 2 x 2 x 2 x 2 *factorial* design. A score reflecting the effectiveness of approval as a reinforcer was the dependent variable.

Four continuous independent variables, representing *S*s' characteristic Attention, Approval, and Achievement orientations, and their Ages, were related to *changes* in this same dependent variable caused by Deprivation, in a *regression* analysis. (Each *S* having served under both Deprivation and Nondeprivation conditions, the *change* in the effectiveness of approval assignable to Deprivation can be taken as the difference between the scores gathered from each *S* under these two conditions.) The detailed descriptions of these variables follow.

FACTORIAL DESIGN

Independent Treatment Variables

Social deprivation. On one of the two occasions when each *S* came to play, he was subjected to a 20-minute period of social isolation before the game (*Deprivation*).

[1] This study, which was carried out when the senior author was at the University of Chicago, was facilitated by a grant-in-aid to him from the Social Science Research Committee of that institution. The authors are grateful to Mrs. Chaya Roth for her very helpful assistance.

[2] At the time of this study the junior author was a Public Health Service Predoctoral Research Fellow of the National Institute of Mental Health at the University of Chicago.

[3] According to the logic used here, many apparent complexities of social contact and dependence may be referred to the assumption that stimulus events provided by the positive social responses of people (e.g., attention, approval, affection) function essentially as positive *reinforcers*. That is, the supply of these stimuli contingent upon various behaviors can systematically increase the outputs of those behaviors (6).

On this occasion, when *S* (escorted by *E*) arrived at the experimental room, a familiar adult would appear to announce that the toy was "broken" and was "being fixed." *S* was given plausible assurance that the toy would be repaired "in a few minutes," and that since *E* wanted him not to miss his turn with the game, he might as well wait in the experimental room, since the adjoining "repair" room was clearly crowded. *S* was then left to wait alone without playthings in the experimental room with the door closed (or slightly ajar if he was more comfortable that way), while both adults retired to "repair" the toy. (During this process, occasional appropriate noises were emitted to indicate that repairs were progressing.) This procedure was carefully structured so as not to imply sanctions, and appeared adequate to guard against the possibility that *S* might feel rejected. If *S* poked his head out and asked to leave during the 20-minute isolation period, he was warmly reassured that the toy was "almost fixed," and he returned to the room. (This rarely occurred.) After 20 minutes, *E* returned with the "repaired" toy, and the game was played in the usual fashion. Under the control condition (*Nondeprivation*), *S* began the game immediately upon his arrival at the experimental room from his nursery group.

Sequence. Each *S* played three games with a given *E*, the last two separated by at least one day. The first of these, a Nondeprivation condition, was to introduce *S* to the game and to *E*, and to give him some familiarity with the experimental setting. (Data from this adaptation session are not included in the analysis.) For half of the *S*s, the second session was preceded by the social isolation condition, and the third was not; for the other half, this order was reversed.

Sex of experimenter. Half of the *S*s in the sample played their three games with a male *E*, and the other half with a female *E*. Both *E*s were graduate students in their early twenties. In dress and appearance, they were not unlike typical adults in the nursery school environment.

Sex of child. Each *E* served with four boys and four girls under each sequence.

Dependent Variable

After the experimental treatment of Deprivation or Nondeprivation, the task for *S* was to drop marbles into a toy while *E* looked on. The toy was a wooden, angular box, having two holes at its forward top into which marbles could be dropped, and an open tray at its base where they returned. The tray contained six marbles, identical in size but different in color, which could be used repeatedly. The experimental room contained two chairs, a couch, and a low table upon which the toy rested. Each *S* was introduced to this setting by *E* as if to a game. *E* and *S* sat before the toy, and *S* was told that he could drop marbles down either hole, but that he must drop only one at a time. *E* then observed *S*'s play for a "baseline" period, during which reinforcers were *not* dispensed, until *S*'s responses appeared relatively stable (typically by four minutes). Meanwhile, *E* responded to *S*'s comments and questions in a friendly but brief manner. Without pause, the 10-minute test of reinforcer effectiveness began, as a continuation of the baseline period. *E* proceeded to dispense the reinforcer according to a schedule immediately after *S* dropped marbles into the hole preferred *least* during the last minute of the baseline period (the *correct* response).

Reinforcement. The reinforcer, designed to appeal primarily to the concept of approval, consisted most frequently of the word "Good!" and less frequently of such phrases as "Hm-hmm," "Good one!" "That's a good one!" and "Fine!" These reinforcers were delivered by *E* in a casual manner, according to a schedule incorporating four successive fixed ratios: 1. At the beginning of the first minute of reinforced play, every correct response was reinforced until 10 consecutive correct responses had been made. 2. Subsequently, five reinforcers were delivered, one for every *second* correct response. 3. Then, five reinforcers were delivered, one for every *third* correct response. 4. Finally, a reinforcer was delivered for every *fifth* correct response until the end of the 10 minutes of reinforced play. This schedule was qualified by three conventions: 1. Every correct response that immediately followed an incorrect response was reinforced. 2. If *S* made five consecutive incorrect responses, *E* began the schedule again, halving the number of reinforcers called for at each ratio level, until the reinforcement ratio last employed was again in effect. 3. If *S* made 20 successive incorrect responses, *E* said, "You can put the marbles in *either* hole."

Reinforcer effectiveness score. The numbers of marbles dropped in the correct and incorrect holes during each minute of play were recorded during the game (which immediately followed the deprivation or control treatment) by an observer in another room. These basic data generated from one to five pairs of scores for the baseline level period, and 10 pairs of scores for the reinforcer effectiveness test period. For each minute of play, the *relative frequency* of a correct response was the ratio of correct responses to total (i.e., correct + incorrect) responses. The *median* of these 10 test ratios was employed as a single representative score reflecting reinforcer effectiveness *uncorrected* for baseline level. The baseline level was taken to be the ratio from the final minute of the baseline period, when no reinforcers were dispensed. (This ratio was assumed to be maximally free from possible "settling down" variations expected during the initial minutes of play, and hence to display the greatest continuity with the following 10-minute reinforcer period.) The corrected effectiveness of the social reinforcer was assessed by taking the difference between the median of the 10 reinforcer test ratios and the baseline ratio. This *reinforcer effectiveness score* represents the gain in relative frequency of the correct response attributable to the social reinforcer of approval.

REGRESSION DESIGN

Independent Variables Reflecting Individual Differences

Five experienced nursery teachers were asked to rate each of the *S*s in their groups for the degree to which *S* characteristically displayed behaviors apparently aimed at securing attention or approval, or indicating a high achievement standard. (These characteristics were thought to represent possible determinants of the reinforcing effectiveness of approval for *S*s.) It was not practicable to ascertain the reliabilities of the ratings. The instructions to the raters stated that the three

scales were conceptually independent. All children in each class were rated on each scale at one time; ratings were made on continuous 100 cm. lines subdivided for reference into five areas. These ranged from "completely indifferent" through "slight," "moderate," "considerable," and "extreme" orientations. Each teacher's ratings were then transformed into a set of standard scores, for each scale separately. (These standard scores were constructed from ratings of *all* the children in the nursery school, of whom the experimental *S*s represent a sample.) The descriptions of these dimensions for the rater were as follows:

Seeks attention or notice from teacher: The child seeks attention or notice from the teacher, in settings where such attention or notice does not imply approval; e.g., "Watch me open this book," and "Look at me" (when the child is not doing something for which he could be praised). The child may deviate, cry, or whine to get the teacher to attend to him.

Seeks approval from the teacher: The child seeks to gain approval or recognition (praise) from the teacher either by direct request (e.g., "Am I not a nice boy?") or by performing in a praiseworthy manner (good behavior, achievement, and the like).

Has high achievement standards: The child who has set high standards for himself works for the sake of excelling at tasks (for which he may or may not receive recognition), is autonomously involved in his achievements, and would attempt new tasks and problems for the chance of solving them successfully.

Age. The age of each *S* was recorded, yielding a set of scores ranging from 46 to 63 months.

Dependent Variable

Deprivation score. The *increase* in the reinforcer effectiveness score from the Nondeprivation condition to the Deprivation condition serves as the dependent variable in the regression analysis. It is computed for each *S* by subtracting the reinforcer effectiveness score displayed under Nondeprivation from that displayed under Deprivation. This *Deprivation score* reflects for each *S* the gain in relative frequency of the reinforced response that may be attributed to the condition of social deprivation.

SAMPLE

Thirty-two middle-class *S*s, ranging in age from 3–10 to 5–3, were selected at random from play groups of a university nursery school. Two of the 16 girls were Negro, the rest of the girls and all 16 boys were white. Otherwise, the primary criterion used in selecting any *S* was his willingness to participate in the "game" on the three occasions comprising the experiment.

RESULTS

Approval as a Reinforcer

A necessary preliminary question is whether the reinforcer dispensed was effective as such, independent of the Deprivation condition. The eight mean reinforcer effectiveness scores under the Nondeprivation conditions were tested in a 2 x 2 x 2 factorial analysis of variance (Sequence × Sex *E* × Sex *S*), and could be considered as drawn from the same population. The hypothesis then was tested that the mean reinforcer effectiveness score under the combined Nondeprivation conditions was zero, employing an error term based upon those scores and a one-tail *t* test. That mean, .10, was found to be reliably greater than zero at $p < .025$, supporting an affirmative conclusion.

TABLE 1
REINFORCER EFFECTIVENESS SCORE MEANS

Sex of *E* and *S*	*N*	Condition	
		Nondeprivation	Deprivation
Male *E*			
Boys	8	.13	.06
Girls	8	.17	.25
Boys & Girls	16	.15	.16
Female *E*			
Boys	8	−.04	.34
Girls	8	.15	.18
Boys & Girls	16	.06	.26
All children	32	.10	.21

Note:—Each *S* contributes both to Nondeprivation and Deprivation means.

Factorial Analysis

The means of the reinforcer effectiveness scores for the experimental conditions of Deprivation, Sex of *E*, and Sex of *S*, the conditions which produced reliable effects, are presented in Table 1. The Sequence variable did not enter into any reliable relationships, and was collapsed for this report. Each *S* yielded two scores, one under Nondeprivation and one under Deprivation. Table 2 presents a summary of the analysis of variance that provides the basis for evaluating differences among the group means. The statistical model employed in the analysis was an extension of Lindquist's Type III (14, p. 281).

The Deprivation variable contributed reliably to the variance in reinforcer effectiveness scores at $p < .025$. The mean reinforcer effectiveness score under Deprivation (.21) was twice that under Nondeprivation (.10). However, the effect under Deprivation varied as a function of which *E* delivered the reinforcer, according to the reliable first-order interaction effect between the variables Deprivation and

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE OF REINFORCER EFFECTIVENESS SCORES

Source of Variation	*df*	Mean square
Between subjects	31	
Sequence (Seq)	1	1,234
Sex *S*	1	736
Sex *E*	1	8
Seq × Sex *S*	1	284
Seq × Sex *E*	1	2
Sex *S* × Sex *E*	1	395
Seq × Sex *S* × Sex *E*	1	1,324
Residual	24	1,531
Within subjects	32	
Deprivation (Dep)	1	1,795**
Dep × Seq	1	11
Dep × Sex *S*	1	479
Dep × Sex *E*	1	1,671*
Dep × Seq × Sex *S*	1	570
Dep × Seq × Sex *E*	1	820
Dep × Sex *S* × Sex *E*	1	2,438***
Dep × Seq × Sex *S* × Sex *E*	1	62
Residual	24	297

* Yields *F* ratio significant at $p < .05$.
** Yields *F* ratio significant at $p < .025$.
*** Yields *F* ratio significant at $p < .01$.

Sex of *E* ($p < .05$). A reliable second-order interaction effect at $p < .01$ provides a final qualification by sex membership of the child and adult on the Deprivation effect, and clarifies the nature of the Deprivation × Sex of *E* interaction effect. Social deprivation enhances the effectiveness of approval as a reinforcer essentially as a cross-sex function: the reinforcer dispensed by each *E* appears on the average more effective after Deprivation only with *S*s of the opposite sex, with this difference reliable solely for boys with the female *E*. With *S*s of the same sex, the effect of Deprivation on the reinforcer appears negligible (and, in the case of boys with the male *E*, apparently in the reverse direction). Appropriate one-tail *t* tests indicate that the reinforcing effectiveness of the male *E* with boys under Deprivation is questionable ($p < .15$); in all other conditions, the male *E* has clear reinforcing value at $p < .05$. The female *E* appears to have no reinforcing value for boys under Nondeprivation, but is reliably increased by Deprivation ($p < .01$) to greater than zero effectiveness ($p < .05$). With girls, the reinforcing value of the female *E* is questionable ($p < .15$ under Nondeprivation, $p < .10$ under Deprivation).

Regression Analysis

It will be recalled that the *Deprivation score* reflects for each *S* the increase in reinforcer effectiveness attributable to the condition of Deprivation. The four continuous independent variables of Age and Attention, Approval, and Achievement orientations were entered into a multiple regression analysis, with the Deprivation score as the dependent variable.[4] (This analysis requires the usual assumptions about the simple regressions involved (9, p. 170), as well as the assumption of homogeneous linear regressions for the eight cells of the 2 x 2 x 2 factorial in the case of every pair of those five variables.) The analysis demonstrates the independent contribution of each of these variables to the Deprivation score, and for the present problem focused only on the identity of the partial regression coefficients found to be reliably different from zero. The Achievement variable was eliminated because its partial regression coefficient was not reliably different from zero. The analysis then yielded a corrected (9, pp. 433–34) multiple correlation coefficient of .82 (20 *df*, $p < .001$).[5] The partial regression coefficients of Age, Approval, and Attention are reliably distinct from zero (at $p < .001$, $p < .01$, and $p < .01$ respectively, for 20 *df*). (These were evaluated by calculating *t* as the ratio of each coefficient to its standard error.) The regression equation was:

Deprivation score (predicted) = 3 (Age) + 1 (Approval) − 1 (Attention) − K. (The

[4] This analysis employed the residual moments (variances and covariances) of these scores, which resulted from isolation of the moments generated by the classifications of the 2 x 2 x 2 factorial design (Sequence, Sex of *S* and Sex of *E*), with the Deprivation and Nondeprivation conditions collapsed to form the Deprivation score. Thus the analysis is essentially independent of the levels determined by those experimental classifications, and represents a distinct further analysis of social deprivation, reflected now in the dependent variable.

[5] Degrees of freedom for the multiple correlation coefficient and for each partial regression coefficient were computed as the total number of *S*s (32) less the number of effects generated by the factorial classifications of the Deprivation score (seven: three main effects, three first-order interactions, one second-order interaction) less the number of variables involved in the multiple regression (five including Achievement, four excluding it) less one.

TABLE 3
RESIDUAL INTERCORRELATIONS OF AGE, ATTENTION, APPROVAL, ACHIEVEMENT, AND DEPRIVATION SCORES[a]

	Attention	Approval	Achievement	Deprivation
Age	.44*	.49*	.32	.75***
Attention		.36	.41*	.11
Approval			.34	.58**
Achievement				.20

[a] Each r, determined from the residual moments of analyses of covariance, has 23 df. Significance levels are denoted as follows: *$p < .05$; **$p < .01$; ***$p < .001$. An r of .34 is reliable at $p < .10$.

final term of this equation is a constant specific to the data of this experimental case.)

The 10 simple (residual) intercorrelations among the four independent variables and Deprivation are presented in Table 3. The correlations of the Deprivation score with Age (+.75, $p < .001$) and Approval (+.58, $p < .01$) are seen to be reliable and positive, while the correlation of Deprivation with Attention (+ .11) is not reliably different from zero. In addition, the low correlations of Attention with Age (+.44, $p < .05$) and with Approval (+.36, $p < .08$) are positive and similar in magnitude and reliability. Together with the reliable negative partial regression coefficient found for the Attention variable ($p < .01$) this correlation pattern can only indicate that Attention functions as a *suppressor* variable (15), to subtract from the variables Age and Approval a portion of the variance in each which is not related to the dependent variable of Deprivation. In this way, the regression coefficients of Age and Approval on Deprivation would be enhanced, the former somewhat more than the latter. Hence, the negative partial regression coefficient for Attention does not represent a negative relationship to Deprivation.

DISCUSSION

As in previous studies, the reinforcing character of verbal approval has been demonstrated. It has been assumed here that approval had acquired reinforcing value through *S*'s previous history of conditioning. It might be expected therefore (*a*) that approval would function as a more effective reinforcer than would (for example) a comparable noise emitted by a machine, and (*b*) that such a noise would be less affected by a preceding period of social isolation. (A demonstration of both of these points clearly would strengthen the conclusions to be drawn in this study.) Similarly, verbal approval from adults has been assumed to represent the social reinforcers controlling the purely social initiations made by children to adults, and the generality of the conclusions must necessarily rest on this assumption.

The effect of social deprivation upon approval as a reinforcer. The data appear to support the conclusion that the effectiveness of a social reinforcer may be increased by the deprivation of social reinforcers generally. This conclusion is immediately qualified by the observation that it holds differentially according to Sex of *E* and Sex of *S*. It is clear that the conditions under which the effect obtains remain further to be delimited, since sex membership in itself hardly explains the matter.

Further, it is important to attend to the postulated equivalence of social isolation and the deprivation of social reinforcers. Clearly, isolation implies the absence of all agents of social reinforcement. Even so, it would be difficult to deny that isolation may imply other characteristics significant to the behavior observed here. For example, isolation might be expected to increase the vigilance of the child, to set up a condition of frustration or annoyance, or to give an impression of reproof or disapproval, and any of these concepts might yield predictions in accord with the data. However, other aspects of the data at hand seem to support the equation of isolation to the deprivation of social reinforcers.

One such aspect is afforded by the (partial) regression of the children's approval and attention orientations on the Deprivation score, which reflects the effect of social deprivation on the reinforcing power of approval in the game. It was found that social deprivation increased reliably the reinforcing power of adult approval for *S*s as a positive function of the degree to which they typically sought approval from the nursery teacher (and presumably from similar adults).

Let us assume for the moment (*a*) that social deprivation represents the unavailability of both approval and attention (among other social reinforcers), and (*b*) that those reinforcers are not equal alternatives for *S*, i.e., that there is for each *S* a reinforcer hierarchy such that some one reinforcer has primary

appeal for him. Then those *S*s who typically seek approval would have been most susceptible to the reinforcer supplied in the game after isolation; but those *S*s who typically seek attention may have found that *E*, through his constant attentive interest, was dispensing a uniform amount of attention independent of their behavior in the game. Since the reinforcer offered contingent upon a correct response would be essentially irrelevant to these attention-seeking *S*s, then no gain in control of that response in the game would follow from the condition of social deprivation. Rather, the *S*s whose attention-seeking tendencies increased as a function of isolation may have employed other behaviors aimed at securing more attention from *E* than the standard amount offered in the experimental setting. And these behaviors, unscored in this study, would have competed with the game-playing response for approval. On this basis, it would be expected for this type of experimental setting that the approval-seeking tendencies of *S*s would relate positively, and their attention-seeking tendencies would relate negatively, or not at all, to the effects of social deprivation upon the reinforcing value of approval. This pattern is in accordance with the conclusions yielded by the multiple regression analysis.

In other words, the results are consistent with the premise that social deprivation sensitizes primarily those social stimuli which are in fact reinforcers for the child deprived. This reasoning could lead to a more precise statement of the experimental conclusion: the effectiveness of a social reinforcer may be increased by its own deprivation. This suggestion might be checked in a design that deprived *S*s of particular social reinforcers (e.g., the one most important for them) but not of all others. If replicated in a sufficient variety of settings, the suggestion could perhaps provide a criterion for a conditioned drive convention, insofar as it specifies an operation of deprivation for a presumably conditioned reinforcer paralleling the criterion of a standard drive operation of deprivation for primary reinforcers. The validity of such a convention would be increased by demonstrating the responsiveness of specific social reinforcers to the *satiation* operations against which the effect of deprivation typically would be reflected; this is the object of a companion study.[6]

Furthermore, a similar effect was found in a previous study (5) of presumably related phenomena involving different *E*s from those of this study and *S*s in the same age range. In that experiment, individual children engaged in easel painting could emit attention-seeking responses to an adult under conditions of the apparent low and high availability of the adult as a source of reinforcers. Similar control was exercised over the sex membership of the *S*s and the two adults of each sex. The same cross-sex pattern of behavior for a social reinforcer class (in that case, attention) was observed, the low availability condition effecting an increase in these behaviors relative to high availability in a manner very similar to the effect of the Deprivation condition in this study. Hence, it appears reasonable to emphasize what appears basic to both conditions, the relative availability of, or access to, social reinforcers. On this basis, social deprivation could be conceived as an operation that need not require isolation for its implementation. That is, the relative availability of social reinforcers might be manipulated in a variety of ways, some involving the actual presence of an agent of reinforcement, as in the low adult availability condition of (5).

Finally, it may be important to consider the possibility that some *S*s might have accidentally experienced mild reproof, rejection, boredom, or the like at *E*'s hands. If such a class of variables were to increase sensitivity to social stimuli like approval or attention (cf., e.g., 10, 16), then it would show an increased reinforcing power, but not because of

[6] In the companion study to be published entitled, "The Effects of Deprivation and Satiation on Behaviors for Social Reinforcers," each of three independent groups of 34 *S*s, aged 6-6 to 9-0, served with a female *E* under similar Deprivation and Nondeprivation conditions as those of this experiment, and a Satiation (for approval and social contact) condition. The mean reinforcer effectiveness score was highest under Deprivation, intermediate under the Nondeprivation condition, and lowest under Satiation, unqualified by any interaction with sex of *S*s. And since the Satiation condition represents a logical extension of a dimension best characterized by a deprivation logic, the mean under this condition rank-orders consistently with the logic. These results give further support to the contention that social isolation represents essentially a deprivation of social reinforcers.

the operation of deprivation (or of *E*'s low availability) as such. Thoughtful care was taken to guard against these possibilities. The setting was designed to be natural and permissive, and *S*s were handled with warmth and reassurance. The possibility that some *S*s might have felt reproved, bored, rejected, etc. cannot be ruled out, but the generally good-natured responses of the *S*s to isolation make these effects seem unlikely.

While these arguments seem to support the idea that social isolation is readily conceptualized as a deprivation of social reinforcers, considerable conceptual caution seems nevertheless to be in order. Probably the conclusion is best considered as tentatively supported pending data from more extensive parametric investigations of these variables.

Age and the effects of social deprivation. A final characteristic of deprivation as defined here is its greater effectiveness upon the older *S*s, evidenced by the positive relationship of Age with the Deprivation score. It is the older children in the sample for whom the social reinforcers of approval and attention are most significant, according to the reliable positive correlations with Age of the Approval and Attention ratings. Clearly, even in the narrow age ranged studied, it is the older children who may be expected to have the longer and better integrated histories of conditioning of these reinforcers. And it is not unreasonable to assume that children in this age range will have as yet only the beginnings of a history of conditioning of achievement as a reinforcer, a conclusion in accord with the zero-order relationship of the Achievement rating to the Deprivation score.

Sex and the effects of social deprivation. The behavior pattern of the boys with the female *E* under the two levels of social deprivation clearly is the primary determinant of the reliable second-order interaction of social deprivation with Sex of *S* and Sex of *E*. This interaction effect might be referred to a number of the rationales summarized and examined in (5) (the most prominent probably being Freud's Oedipus theory [4]). However, cautious interpretation is in order. The intent of the experiment was of course to exercise some control over the sex of the reinforcer-dispensing adult relative to the sex of the child; yet it must be observed that only one *E* was used to represent each sex. These two persons surely differed with respect to a great number of criteria, many of them of considerable theoretical pertinence. Thus, the reliable interaction effects in which the identity of the *E*s figure might be referred as reasonably to some of these possible differences as to sex. Only a wide replication of these interactions with suitably chosen *E*s can resolve this ambiguity. Still, the very similar cross-sex effect found in the Availability study cited (5) suggests that it is the sex patterns of *S*s and *E*s that are at issue.

SUMMARY

On the assumption that approval is representative of the reinforcers controlling the purely social initiations made by young children to adults, the verbal approval of an adult (*E*) was made contingent upon one response in a two-response game for 32 nursery school children (*S*s). The change in relative response frequency from a baseline level following introduction of adult approval, reflected the degree to which approval was reinforcing for an *S*. Adult approval was found to have definite reinforcing value for the *S*s ($p < .025$). With control exercised over sex of child, sex of adult, and sequence of conditions, the mean effectiveness of adult approval as a reinforcer for the *S*s was enhanced reliably, by a preceding period of 20 minutes of social isolation (conceived to be social deprivation), relative to its effectiveness for the same *S*s when not isolated ($p < .025$); and this result was qualified by the fact that the reinforcer dispensed by each *E* appeared more effective after social deprivation only with *S*s of the opposite sex, with this difference reliable alone for boys tested by the female (rather than the male) *E*. The increase in effectiveness of adult approval as a reinforcer produced in the children by social isolation was shown to be a positive function of the degree to which they characteristically sought adult approval in other settings ($p < .01$), and of age ($p < .001$), in the 46- to 63-month range studied.

Because social isolation excluded access to persons who might provide social reinforcers, its effect was referred to a concept of social deprivation, defined as the unavailability of social reinforcers. Thus it appears that the effectiveness of a representative social rein-

forcer may be enhanced by an operation of deprivation, as is the case for the reinforcers of primary appetitive drives like hunger and thirst. The effectiveness of this reinforcer appears enhanced particularly for *S*s who typically seek it.

REFERENCES

1. BOWLBY, J. Maternal care and mental health. *Bull. World Hlth Org. Monogr.*, 1951, **3**, 355–533.
2. BUTLER, R. A. The effect of deprivation of visual incentives on visual exploration motivation in monkeys. *J. comp. physiol. Psychol.*, 1957, **50**, 177–179.
3. CHASE, LUCILLE. Motivation of young children: An experimental study of the influence of certain types of external incentives upon the performance of a task. *Univ. Iowa Stud. Child Welf.*, 1932, **5**, No. 3.
4. FREUD, S. The passing of the Oedipus complex (1924). In *Collected papers*, Vol. 2. London: Hogarth, 1949. Pp. 269–276.
5. GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Monogr. Soc. Res. Child Develpm.*, 1954, **19**, No. 2 (Serial No. 59).
6. GEWIRTZ, J. L. A program of research on the dimensions and antecedents of emotional dependence. *Child Develpm.*, 1956, **27**, 205–221.
7. GOLDFARB, W. Effects of psychological deprivation in infancy and subsequent stimulation. *Amer. J. Psychiat.*, 1945, **102**, 18–33.
8. GREENSPOON, J. The reinforcing effect of two spoken words on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
9. GUILFORD, J. P. *Fundamental statistics in psychology and education*. (2nd ed.) New York: McGraw-Hill, 1950.
10. HARTUP, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. Unpublished doctoral dissertation, Harvard Univer., 1955.
11. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
12. HURLOCK, ELIZABETH B. The value of praise and reproof as incentives for children. *Arch. Psychol.*, N. Y., 1924, **11**, No. 71, 5–78.
13. KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology*. New York: Appleton-Century Crofts, 1950.
14. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
15. LUBIN, A. Some formulae for use with suppressor variables. *Educ. psychol. Measmt*, 1957, **17**, 286–296.
16. ORTMEYER, D. H. Drive generalization from hunger to fear. *Dissertation Abstr.*, 1955, **15**, 1257–1258. (Abstract)
17. RIBBLE, MARGARET. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 621–651.
18. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476–483.
19. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
20. SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. In Anna Freud *et al.* (Eds.), *The psychoanalytic study of the child*. Vol. 1. New York: Int. Universities Press, 1945. Pp. 53–74.
21. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.
22. WOLF, THETA S. The effect of praise and competition on the persisting behavior of kindergarten children. *Univer. of Minnesota Inst. Child Welf. Monogr. Ser.*, 1938.

Received March 27, 1957.

PREFERENCE FOR DELAYED REINFORCEMENT: AN EXPERIMENTAL STUDY OF A CULTURAL OBSERVATION[1]

WALTER MISCHEL

University of Colorado

THIS study is, in part, an attempt to validate, experimentally, an observation about cultural differences formulated during the course of a "culture and personality" field research in Trinidad, British West Indies. It may be of interest, therefore, not only for its specific findings but also as an example of experimentation in the cultural field.

On the basis of anthropological observation, a major personality difference between the Negro and East Indian ethnic groups of the island of Trinidad was suggested. This difference, as expressed by numerous informants, is that the Negroes are impulsive, indulge themselves, settle for next to nothing if they can get it right away, do not work or wait for bigger things in the future but, instead, prefer smaller gains immediately. In contrast to this, the Indian is said to deprive himself and to be willing and able to postpone immediate gain and pleasure for the sake of obtaining greater rewards and returns in the future. In effect, when given a choice, the Negro is said to be characterized by preference for relatively smaller, immediate reinforcements, whereas the Indian is said to prefer larger, delayed reinforcements.

Judgments centering about such a distinction of differential preference or tolerance for delayed reinforcement are, of course, also quite prevalent within our own culture. The widely observed preferences of children for immediate rewards has been commented on (2, 3, 5) and is reflected in some theoretical formulations, e.g., the "pleasure principle" and the "reality principle" (4, 6). In clinical applications, the inability to postpone immediate gratification for the sake of delayed rewards is often considered an important factor in immaturity, maladjustment, and "psychopathy." Mowrer and Ullman (10), for example, consider the delay of reinforcements to be a critical factor in the development of neurotic and criminal behavior.

Although the research on delay of reinforcement in human conditioning has been vast (Wolfle's studies [14, 15] are representative), there has been comparatively little experimentation with human subjects on variables other than the effect of the time interval itself on learning rate or object preference (e.g., 1, 7, 8, 11, 13) and outside the conditioning paradigm. A notable exception is Mahrer's (9) attempt to apply Rotter's expectancy construct (12). Mahrer was able to make successful predictions about children's choices for immediate versus delayed reinforcements as a function of differential expectancies for obtaining positive reinforcement following delay. That is, the expectancy that reinforcement would follow from the social agent even after time delay was shown to be an important variable. In common sense terms, this kind of behavior may be thought of as "trust" or the belief that the agent promising delayed reinforcement will actually supply it. The emphasis in Mahrer's study, and in the present research, is on the expectancy for reinforcement to follow as a function of making a particular choice with a particular social agent in a given situation, rather than on the time interval itself.

One may assume that expectancies for reinforcement to follow from certain social agents, in spite of time delay, are learned in a manner fundamentally similar to the learning of any social behaviors, and are governed by similar principles of generalization. It seems probable, then, that a child's expectancies that delayed reinforcement will issue from new adult social agents are related to his past experiences in which promised reinforcement followed delay from such major adult sources of learning as the parental figures within the home. In a situation in which a child is confronted with a new male social agent, we would anticipate that his expectancies for promised reinforce-

[1] This study was conducted while the author collaborated with Frances Mischel in an interdisciplinary project which was sponsored by the Wenner-Grenn Foundation for Anthropological Research. A report based on portions of this paper was read at the convention of the Rocky Mountain Psychological Association, April, 1957.

ment to follow from that social agent, in spite of time delay, are related to past experiences of this kind with the father or father figure in the home. Similarly, we would anticipate that the child's expectancies that promised reinforcement will follow, in spite of delay, from female social agents are related to past experiences of this kind with the mother or mother figure in the home. (Since in the present experiment the social agent (*E*) is male, our discussion and prediction center about expectancies for delayed reinforcement from male social agents.) We would further expect children who have had less of such experience with a male social agent in the home to show less "trust," or lower expectancy for reinforcement to follow in spite of time delay when promised by a new male social agent, than children who have had more of such experience. Presumably, a child whose father is absent in the home has had more limited experiences in which promised reinforcement follows in spite of time delay from a male social agent. We anticipate that such a child would have lower expectancy for delayed reinforcement from a male social agent than a child whose father is present within the home. Specifically, we would expect the "fatherless" child to show less preference for a larger but delayed reinforcement and more preference for a smaller but immediate reinforcement than the child whose father is present within the home.

The above prediction seemed, on the basis of tentative observation, to apply differentially to the two cultural groups we are considering; namely, examination of the family structures of the Negro and Indian groups within the particular culture and locale suggested that the Negro family seemed more frequently characterized by the absence of a temporally stable adult male figure in the home than was the Indian family. The presence or absence of a father within the family constellation was selected as a key variable influencing choice behavior with respect to delayed or immediate reinforcement and the aim of the experiment was: (*a*) to test the hypothesis of differential preference for larger, delayed as opposed to smaller, immediate reinforcement of the two ethnic groups, it being predicted that the Indian group would show greater preference for larger, delayed reinforcement than the Negro group; and (*b*) to test the hypothesis of differential preference for larger, delayed as opposed to smaller, immediate reinforcement, issuing from a male social agent, of children whose fathers are present or absent within the home, it being predicted that children having a father or father figure present within the home would prefer the larger, delayed reinforcement more than children lacking such a figure. Further, since age and socioeconomic level seemed potentially relevant variables for this kind of choice behavior, this information was included in the data collected, with the intent of relating it to the choice.

METHOD

Subjects

The subjects (*Ss*) included 53 male and female children between the ages of 7 and 9 in the elementary section of a rural Trinidad school. The school was located in the southern part of the island and seemed representative of rural schools within the colony. Thirty-five of the children were Negro and eighteen East Indian. The sex ratio was 29 males to 24 females. The children of the two ethnic groups attended the same large section or class and it was thus possible to administer the experimental test to the members of both ethnic groups at the same time.

The Reinforcements

Pre-experimental sessions were required to select two reinforcements of use for the experimental situation. Towards this end, 15 Negro and Indian boys and girls, representing the above ages, and selected from another rural Trinidad school, were seen in individual sessions and their preferences for various specific reinforcements were elicited. As a result, two reinforcements (both candy, but varying markedly in size, price, and packaging, i.e., a one-cent and a ten-cent candy) were selected. These met the desired requirements inasmuch as the larger reinforcement was uniformly preferred in a straight choice situation ("which *one* of these two would you like to take"), but when the choice was "you can have this one (the smaller) today *or* this one (the larger) in one week" approximately 50 per cent of the group chose the former and approximately 50 per cent the latter.

Task and Measures

The *Ss*' task was to fill out a simple questionnaire. All the *Ss* were sufficiently literate for this. The questionnaire items of experimental interest were ethnic group, age, socioeconomic status, and presence or absence of the father within the household. Ethnic group was measured by *S*'s checking either "Creole" or "Indian," and age by a "fill in" response. The measure of socioeconomic status was quite gross, and *S* merely checked "galvanized" (tin roof, comparatively expensive and generally a sign of relatively high socioeconomic status in the rural areas) or "carrat" (thatched leaf roof, inexpensive, and generally a sign of

low socioeconomic status), in response to an item concerning the kind of house he lives in. This measure provided a crude dichotomization into relatively "high" and relatively "low" socioeconomic status. "Presence" or "absence" of the father in the home was measured by the response "yes" or "no" to the questions "does your father live at home with you," and "do your mother *and* father live at home with you." One *S* reported that father lived at home, mother lived at home, but that mother and father did not live at home and was eliminated from the analysis for this inconsistency. All but three *S*s reported "yes" to the question "does your mother live at home with you." These three exceptions lived with their grandmothers and because of this deviation were omitted from the statistical analyses.

An informal check on the accuracy of the children's responses was attempted with the headmaster who was familiar with the home situations of most of the *S*s. This check was undertaken on a sample of twenty of the protocols and indicated that, to the headmaster's best knowledge, the responses corresponded with his own information about the children.

Procedure

The experimenter (*E*) was introduced as an American interested in gathering information on children in the local schools of the island. To help with this the *S*s were asked to fill out the questionnaire. When these were completed, *E* expressed his wish to thank the group for their cooperation. He displayed the two kinds of reinforcements and said: "I would like to give each of you a piece of candy but I don't have enough of these (indicating the larger, more preferred reinforcement) with me today. So you can either get this one (indicating the smaller, less preferred reinforcement) right now, today, or, if you want to, you can wait for this one (indicating) which I will bring back next Wednesday (one week delay interval)." To insure clarity, these instructions were repeated in rephrased form and both reinforcements were carefully displayed. The fact that getting the (smaller) candy today precluded getting the (larger) one next week, and vice versa, was stressed. *S*s were asked to indicate their choice by writing "today" (T) or "next week" (W) on their questionnaires. The response made here was the measure of choice of a larger (or more preferred) delayed reinforcement or a smaller (less preferred) immediate reinforcement. *S*s were seated sufficiently far apart from each other to insure reasonably that their choices were made independently in this group setting.

RESULTS AND DISCUSSION

The data from the present experiment were analyzed in terms of differential preference for smaller, immediate reinforcement versus larger, delayed reinforcement in relationship to ethnic group, presence or absence of the father within the home, age, and socioeconomic status. All statistical analyses were made with the chi-square test. The data used for relating the experimental choice with ethnic group and presence or absence of the father within the home, are given in Table 1.

TABLE 1

RELATIONSHIPS BETWEEN ETHNIC GROUP, PRESENCE OF THE FATHER, AND EXPERIMENTAL CHOICE

Group	*N* Choosing Immediate Reinforcement	*N* Choosing Delayed Reinforcement	χ^2	p
Negro	22 (18.49)[a]	13 (16.51)		
Indian	6 (9.51)	12 (8.49)	4.17 3.06[b]	<.05 <.10
Father present	17 (22.19)	25 (19.81)		
Father absent	11 (5.81)	0 (5.19)	10.13[b]	<.01
Negroes with father present	12 (15.71)	13 (9.29)		
Negroes with father absent	10 (6.29)	0 (3.71)	6.19[b]	<.02

[a] Figures in parentheses are expected frequencies.
[b] Chi square corrected with Yates correction, for continuity.

Comparison of the Negro and Indian groups on the experimental choice (Table 1) yielded a chi square of 4.17 (p between .05 and .02), a significantly larger proportion of the Negro *S*s choosing immediate reinforcement. It should be noted, however, that if the Yates correction is applied (the expected frequencies within two of the cells are between eight and ten), the chi square is reduced to 3.06 having a p value of less than .10. Nevertheless, this difference between the ethnic groups on the experimental choice still seems sufficient to require caution in combining them on other analyses unless it can be shown that members of the two groups are in comparable proportions on the other relevant variables.

The comparison between ethnic groups with respect to presence or absence of the father does not indicate a difference at an acceptable probability level (chi square is 2.56 with $p < .20$ and $> .10$). Some ancillary evidence supporting the trend of differential fatherlessness within the two ethnic groups can be drawn from another source. Data, including the measure of presence or absence of the father within the home, were collected within the same school for older children, of both ethnic groups, who were not included in the present study. If the data from these children ($N = 24$) on presence or absence of the father for the two ethnic groups are combined with data from the *S*s of the present study, a chi square of 5.57 is obtained which is significant

beyond the .02 level, and in the direction expected from cultural observation.

In view of these findings, the relationship between presence or absence of the father and the experimental choice was tested for each of the ethnic groups separately as well as for the total combined sample. As there is only one "fatherless" *S* in the Indian group, such a test is not meaningful for that group. For the Negro group alone, and for the total sample, the comparison between presence or absence of the father and the experimental choice (Table 1) yielded chi squares of 10.13 and 6.19 respectively, which are significant beyond the .01 and .02 levels, and indicate, as predicted, a greater proportion of choices for immediate reinforcement on the part of "fatherless" *S*s.

Before this relationship between presence of the father and preference for delayed, larger reinforcement can be generalized, it will of course be necessary to test other cultural groups in a similar experimental situation. The present research emphasized the relationship only between the "physical" presence of *S*'s father in the home and preference for delayed reinforcement issuing from another male social agent. In view of the obtained significant finding, the relationship between the "psychological" presence of the father (e.g., degree of identification) and this kind of behavior merits further investigation. Such research would seem especially relevant for a further understanding of impulsive behavior and "psychopathy."

The presence or absence of the father may be significant not as, or just as, a causal factor in itself in relation to preference for delayed versus immediate reinforcement, but as an index of the extent to which the parents or the family participate in a delayed reward culture. That is, the presence of the father in the home (in contrast, for example, to a ready abandoning of the home) may, in turn, be related to the degree of family participation in a cultural pattern in which delayed, larger reinforcement is preferred or prized more than immediate, smaller reinforcement. Such participation may have many other impacts upon the child apart from, or in addition to, the paternal presence.

Neither the proportion of Negro and Indian *S*s, nor the proportion of *S*s with father present and absent, was significantly different for the three age groups or for their socioeconomic status groups (both p levels greater than .20). Therefore, tests of the relationship between age and the experimental choice, and between socioeconomic status and the experimental choice, can be made with all *S*s combined. Table 2 shows the number of *S*s in each age group choosing immediate versus delayed reinforcement. The chi-square test applied to this table yields a corrected chi-square value of 7.10 which is significant beyond the .05 level and indicates an increasing proportion of *S*s choosing delayed reinforcement at the older age levels.

TABLE 2
RELATIONSHIP BETWEEN AGE AND EXPERIMENTAL CHOICE

Experimental Choice	Age			χ^2	p
	7	8	9		
N Choosing immediate reinforcement	13 (8.45)[a]	13 (14.27)	2 (5.28)		
N Choosing delayed reinforcement	3 (7.55)	14 (12.73)	8 (4.72)	7.10[b]	<.05

[a] Figures in parentheses are expected frequencies.
[b] Chi square corrected with Yates correction, for continuity.

This relationship between *S*s' age and preference for delayed reinforcement is in accord with nonexperimental discussions of such a relationship to the effect that with increasing maturity comes the increasing ability to delay gratification. The finding is consistent with an interpretation of preference for delayed reinforcement as a learned behavior which is, in part, a function of the expectancy that the promised reinforcement will issue from the social agent in spite of time delay. With increasing age, the potentiality for developing a strong expectancy of this kind increases *if* the individual continues to gain reinforcing experiences within this area, thus building up the relevant expectancies, but not as a function of growing older or biological maturation per se.

Comparison of the "high" versus "low" socioeconomic groups on the experimental choice did not yield a significant difference (p value greater than .20). This lack of established relationship between socioeconomic level and the experimental choice cannot be considered definitive in view of the extreme crudeness of the measure of socioeconomic level employed.

SUMMARY

This experiment tested an observation, made by anthropological field techniques, regarding personality differences between the East Indian and Negro populations of Trinidad, B.W.I. The primary results were significant differences between preference for immediate, smaller versus delayed, larger reinforcement and, first, the presence or absence of the father within the home, and second, age. The over-all findings appear to have implications not only for the two specific ethnic groups studied here but also for further research on relationships between personality variables of a less gross kind and this type of choice behavior within our own culture. If there are relationships between "maturity" or "adjustment" and tolerance, or preference, for larger, delayed as opposed to smaller, immediate reinforcement, then the isolation of major variables related to such preferences seems particularly important.

REFERENCES

1. BRENNER, B. Effect of immediate and delayed praise and blame upon learning and recall. *Teach. Coll. Contrib. Educ.*, 1934, No. 620.
2. COLE, LUELLA, & MORGAN, J. J. B. *Psychology of childhood and adolescence.* New York: Rinehart, 1947.
3. FAEGRE, MARION L., & ANDERSON, J. E. *Child care and training.* Univer. of Minnesota Press, 1937.
4. FREUD, S. *Beyond the pleasure principle.* New York: Boni & Livewright, 1922.
5. HARRIS, D. B. How children learn interests, motives, and attitudes. In N. B. Henry (Ed.), *49th Yearbook, NSSE, Part I, Learning and instruction*, 129–155.
6. HENDRICK, I. *Facts and theories of psychoanalysis.* New York: Knopf, 1948.
7. IRWIN, F. W., ARMITT, F. M., & SIMSON, C. Studies in object preferences: I. The effect of temporal proximity. *J. exper. Psychol.*, 1943, **33**, 64–72.
8. IRWIN, F. W., ORCHINIK, C. W., & WEISS, JOHANNA. Studies in object preferences: The effect of temporal proximity upon adults' preferences. *Amer. J. Psychol.*, 1946, **59**, 458–462.
9. MAHRER, A. R. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, **52**, 101–105.
10. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52**, 61–90.
11. NAGGE, J. W., *Psychology of the child.* New York: Ronald, 1942.
12. ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.
13. SALTZMAN, I. J., Delay of reward and human verbal learning. *J. exp. Psychol.* 1951, **41**, 437–439.
14. WOLFLE, H M. Time factors in conditioning finger withdrawal. *J. genet. Psychol.*, 1930, **4**, 372–378.
15. WOLFLE, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. genet. Psychol.*, 1932, **7**, 80–103.

Received March 25, 1957.

EMPATHY OR BEHAVIORAL TRANSPARENCY?[1]

URIEL G. FOA

The Israel Institute of Applied Social Research and Tel Aviv University

THE term social empathy has been generally used to indicate the activity of guessing another person's perception with regard to the behavior or values of a given individual. For example, A guesses B's perception of C's behavior or values. It is generally understood that A and B are two different individuals. Individual C, however, may be either a third person distinct from A and B, or identical with either A or B, as in a situation involving two people only. It seems quite possible that A might succeed in guessing B's perception of the behavior or values of B, and fail to guess B's perception of the behavior or values of A.

The following four aspects or facets of empathy may therefore be worth considering: (*a*) the guesser, or the individual who guesses the perception of another person; (*b*) the guessee, or the person whose perception is guessed; (*c*) the actor, or the individual whose behavior or values in a particular role are guessed about; and (*d*) the level, which indicates whether the guessing regards the actual behavior of the actor, or the corresponding values.

Most of the previous studies of empathy have been concerned with the guesser and/or the guessee. The actor has been tacitly assumed to be identical with the guessee and, in general, he does not appear explicitly in the design of the experiments reported in the literature. For some investigators, empathy is a personality trait of the guesser that is possessed in different degrees by different subjects. Tests of empathy (2, 5) and studies of the relationship between empathy and other personality traits (8) stem from this conception.

Other students of empathy have held that this variable depends on guesser and guessee alike: if A can predict one aspect of the response of B, he can also predict other aspects of B's response, although he might fail in the prediction of the responses of another person (7).

A third approach is suggested by Tagiuri and others (9). Essentially, it consists in looking at empathy from the vantage point of the guessee rather than from the point of view of the guesser. People, says Tagiuri, are more or less easy to understand; they possess a different degree of transparency. The feelings of transparent people are easier to guess. Empathy scores depend on the guessee.

It is now suggested that success or failure in guessing correctly the perception of somebody else may depend not only on the guesser and/or the guessee but also on the actor. If the actor behaves in an unambiguous manner, it may prove easy to guess the perception that somebody gets of this behavior. Vice versa, if the behavior of the actor is ambiguous, people may get blurred perceptions of it and the guessing of these perceptions may be difficult.

The hypotheses to be tested can be formulated in the following way:

Hypothesis I. Guesses made by the same guesser with regard to different guessees correlate among themselves. For example: if Bill is a good empathizer he will guess correctly the perceptions of other people, no matter who the guessees and the actors are. This is the personality trait approach.

Hypothesis II. Guesses made by different guessers with regard to the same guessee correlate among themselves. This is the transparency approach.

Hypothesis III. Different guesses made by the same guesser with regard to the perceptions of the same guessee correlate among themselves.

Hypothesis IV. Guesses made by different guessers with regard to different guessees correlate among themselves if they all deal with the behavior of the same type of actor. This is the approach introduced in the present paper.

Hypothesis III combines Hypotheses I and II. It is therefore evident that if Hypothesis III is rejected then Hypotheses I and II are also rejected. The acceptance of Hypothesis III, however, does not necessarily prove Hy-

[1] This paper is one in a series of studies on interpersonal relations in industry which have been initiated with partial support from the Ford Foundation, and was read at the 1957 Annual Meeting of the Israel Psychological Association.

potheses I and II. The experiment described in this paper permits the testing of Hypotheses III and IV. These two hypotheses are not mutually exclusive; that is the acceptance of one of them does not lead to the necessary rejection of the other.

METHOD

The Test of Foreman-Worker Relationship (4) was administered to 490 Israeli factory workers and their 51 foremen. This test consists of two sets of pictures of eight situations each. One set is concerned with the behavior of the foreman and the other with the behavior of the worker. Both sets of situations are presented to the foreman and the worker. For each situation they are instructed to choose, out of four given types of behavior, the one which (*a*) ought to occur in an ideal situation, and (*b*) occurs most often in practice in their department. Further, each worker is requested to guess the alternatives chosen by his own foreman, both at the "ought to be" and the actual level. Similarly, each foreman is requested to guess the most frequent (or modal) choice of his workers at both levels.

Let us consider a dyad composed of a worker and his foreman. Each of the two subjects (*S*s) of the dyad gives four sets of responses of eight responses each, thus:

1. Responses regarding the *ideal* behavior of the *foreman*.
2. Responses regarding the *actual* behavior of the *foreman*.
3. Responses regarding the *ideal* behavior of the *worker*.
4. Responses regarding the *actual* behavior of the *worker*.

Further, each *S* guesses the responses of the other.

The *empathy* score for each set of eight responses is computed by counting the number of times the guess of one *S* is the same as the corresponding response of the other. Since there are eight responses in each set, it is evident that the maximal empathy score is eight, and obtained when the *S* has succeeded in guessing correctly all the eight responses of the set and the minimal score zero. Each *S* guesses four sets of responses and thus receives four empathy scores. In a dyad composed of a worker and his foreman there are eight empathy scores, of which four pertain to the worker and four to to the foreman.

A measure of *real similarity* in response for each set of eight pictures was computed by counting the number of times the response of one *S* is the same as the response of the other *S*. A high score indicates that the two *S*s tend to have the same perception of a given type of behavior.

In the population of dyads, the product-moment coefficient of correlation was then computed for each possible pair of empathy scores. Since for each dyad there are eight scores, 28 pairs and their respective coefficients, are possible. Each coefficient indicates the correlation between two empathy scores which differ from one another in at least one of the following elements: (*a*) the *S* with whose behavior the empathy is concerned: either the worker or the foreman; (*b*) the *S* making the guesses and/or the *S* guessed about (in a dyad, when the worker is the guesser, the guessee is the foreman and vice versa); and (*c*) the level of behavior with which the empathy is concerned: either actual or ideal.

When a particular coefficient differs significantly from zero, this suggests that at least one of the elements that are identical in the two correlated scores may be important in the empathic process. On the other hand, when a coefficient approaches zero, this suggests that the elements identical in the two scores are not relevant to empathy in the situation being studied. This kind of reasoning is somewhat similar to factor analysis. In factor analysis, one infers the existence of a common factor when two variables are correlated. Here, each variable is described a priori in terms of certain factors, and the eventual correlation between two variables is attributed to the factors they have in common.

RESULTS

The 28 product-moment correlation coefficients for each possible pair of empathy scores, are reported in Table 1. At the .05 level, coefficients of the correlation between two empathy scores for workers are significant when above .09; coefficients of the correlation between two scores for foremen are significant when above .27. For coefficients between a score of the worker and a score of the foreman the critical value is, of course, somewhere between these two values. An approximation to this value is given by the formula:

$$F = r^2 \Big/ \frac{1 - r^2}{H\,(N-2)},$$ where F is the variance ratio for the .05 probability level; N the number of foremen; and H the harmonic mean of the number of workers per foreman. In the present case, the formula shows that coefficients of correlation above .15 are significant.

In Table 1, there are twelve coefficients between scores referring to the behavior of the same type of actor (upper left and lower right quadrants of the table); eight of them are significant. On the other hand, no significant coefficient is found among the 16 correlations between scores referring to different types of actor (lower left quadrant of the table). For instance: the empathy score of the worker for the norms of the foreman regarding the behavior of the worker correlates at .32 with the empathy score of the foreman regarding the worker's perception of the actual behavior of the worker. Here the actor is the same, but guesser, guessee, and level are different, and yet the correlation is fairly high. On the other hand, the empathy of the workers for

TABLE 1
EIGHT EMPATHIES: INTERCORRELATIONS AND AVERAGE SCORES

Form of Empathy	Correlation Coefficient								
		1	2	3	4	5	6	7	8
Behavior of foreman:									
Worker's guesses of foreman's responses									
Actual level	1	—							
Ideal level	2	.34	—						
Foreman's guesses of worker's responses									
Actual level	3	.26	.19	—					
Ideal level	4	.22	.13	.15	—				
Behavior of worker:									
Worker's guesses of foreman's responses									
Actual level	5	.04	.06	.06	.15	—			
Ideal level	6	.04	.06	−.04	.05	.26	—		
Foreman's guesses of worker's responses									
Actual level	7	.13	.02	.03	−.01	.29	.32	—	
Ideal level	8	.15	−.09	.03	−.03	.06	.27	.34	—
Correlation with real similarity	—	.42	.25	.58	.58	.56	.37	.57	.44
Average empathy score	—	3.44	3.85	3.26	4.24	3.32	4.04	4.08	4.46

the foreman's perception of the foreman's actual behavior correlates at .04 with the empathy of the worker for the foreman's perception of the worker's actual behavior. Here the actor is different, but guesser, guessee, and level are the same in both scores, and the correlation is practically zero. The average coefficient between scores with the same type of actor is .24: the correlation is significant. The average coefficient between scores with a different type of actor is .05, which is below the critical value. These results seem to support the acceptance of Hypothesis IV.

The average correlation between scores with the same type of guesser and guessee is .13. The average correlation between scores with a different type of guesser and guessee is .14. The two coefficients are practically alike and below the critical value, suggesting that Hypothesis III should be rejected. Rejection of Hypothesis III leads, in turn, to the rejection of Hypotheses I and II.

The next to the last row of Table 1 gives the coefficients of correlation between each type of empathy score and the corresponding score of real similarity. These coefficients indicate the extent to which the accuracy of the *S*, in guessing the perception of the other, is related to the fact that both have the same perception. Let us think, for example, of a number of *S*s who always attribute to the other their own perception. Then their guesses would be right when the two perceptions are alike, and wrong when the two perceptions are different. In this case the investigator would find perfect positive correlation between empathy and real similarity scores. In our case, the coefficients of correlation between real similarity score and the corresponding empathy score show that accuracy of guessing is, to some extent, a function of real similarity. The two variables, however, are far from being identical.

In the last row of Table 1 one finds the average score for each type of empathy. The significance of the differences between any two of these averages was checked by the *t* test for correlated pairs and the following results were obtained:

1. The average score at the actual level is always significantly lower than the corresponding score at the ideal level. The significance level is .001 in three cases and .05 in one case.

2. The average scores referring to the empathy of the foreman are better than the corresponding scores referring to the empathy of the worker, with one exception: the score of the foreman regarding the foreman's actual behavior. All these differences are significant; three of them at the .001 level and one at the .05 level.

3. The average scores referring to the behavior of the worker are better than the corresponding scores referring to the behavior of the foreman, again with an exception: the score of the worker regarding the worker's actual behavior. Differences between scores regarding ideal behavior are significant. Those regarding actual behavior are not.

4. The two worst scores are the one of the foreman regarding the foreman's actual behavior and the one of the worker regarding the worker's actual behavior. The difference between each one of these average scores and all the other averages is significant respectively in five and four cases out of six.

DISCUSSION

The findings suggest the following conclusions:

1. The accuracy of A in predicting a certain aspect of the social perception of B is not necessarily related to A's accuracy in predicting another aspect of the perception of B.

2. Empathy scores referring to the perception of the behavior of the same actor are likely to be moderately correlated among themselves.

3. Empathy is partly determined by similarity of perception between guesser and guessee.

It seems possible that the relationship between empathy scores referring to the same type of behavior may be due to the degree of similarity of perception of such behavior. Indeed, any two empathy scores referring to the same level and the same actor are both related to the same similarity score. The real similarity score, for a given level and a given actor, is common to the worker and to his foreman. This suggests that the relationship between empathy scores referring to the same actor might be due to the similarity between the perceptions of the guesser and of the guessee. Indeed, when the common real similarity score is partialled out, the correlation coefficients between empathy scores tend to vanish.

It seems, therefore, that unambiguous behavior is more likely to be perceived in the same manner by different people than ambiguous behavior. If the behavior to be guessed varies considerably in ambiguity, as in the present case, it may influence the relationship between empathy scores. As a consequence, it may be suggested that the type of behavior to be guessed is one of the variables that should be taken into account in the design of studies of social perception and empathy. Failure to control this variable can introduce a source of unexplained variance. Keeping it constant, on the other hand, may help the investigator in determining the portion of the variance of empathy scores which is related to the guesser and/or the guessee.

The comparison of the average empathy scores seems to confirm a previous finding by Nagle (6) that ego-involvement impairs empathy: both foremen and workers are indeed at their worst when guessing the other's perception of the behavior of their respective roles.

The finding that the foreman is a better empathizer than the worker is in agreement with the results reported by Bell and Hall (1) who found that leadership and empathy are related.

On the other hand, our data suggest that values regarding the behavior of the worker are more transparent than the values regarding the behavior of the foreman. Also, they show that it is easier to guess a person's values than it is to guess his perception of actual behavior.

SUMMARY

This is a study of the intercorrelations among eight different kinds of empathy between 490 industrial workers and their respective foremen.

The findings show that measures of empathy referring to the behavior of the same individual are correlated. On the other hand, those referring to the behavior of different individuals are uncorrelated, even when the guesses are made by the same subject and/or with reference to the responses of another given subject.

The present data are thus in disagreement with theories that empathy depends on the personality of the guesser or on the personality of the guessee. They suggest that when people behave in a transparent manner it is easier to guess the perception of this behavior; that is, empathy is related to transparency of behavior. The type of behavior is therefore a variable that should not be disregarded in the study of empathy.

The average scores on the eight kinds of empathy are also discussed. Guessing of the other's perception of the behavior of one's own role is of particular difficulty for both the foreman and the worker. In general, the empathy scores of the foreman are better than the corresponding scores of the worker. Guessing the other's values is easier than guessing his perception of actual behavior.

REFERENCES

1. BELL, G. B., & HALL, H. E. The relationship between leadership and empathy. *J. abnorm. soc. Psychol.*, 1954, **49**, 156–157.
2. BENDER, I. E., & HASTDORF, A. M. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.
3. FOA, U. G. The Foreman-worker interaction: A research design. *Sociometry*, 1955, **18**, 226–244.

4. FOA, U. G. A test of the foreman-worker relationship. *Personnel Psychol.*, 1956, **9**, 469–486.
5. JARRARD, L. E. Empathy: The concept and industrial applications. *Personnel Psychol.*, 1956, **9**, 157–167.
6. NAGLE, B. F. Productivity, employee attitude and supervisor sensitivity, *Personnel Psychol.*, 1954, **7**, 219–233.
7. PATTON, W. M., JR. Studies in industrial empathy: III. A study of supervisory empathy in the textile industry. *J. appl. Psychol.*, 1954, **35**, 285–288.
8. TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **51**, 1–23.
9. TAGIURI, R., KOGAN, N., & BRUNER, J. S. The transparency of interpersonal choice. (Stencilled), Laboratory of Social Relations, Harvard Univer., n.d.

Received April 2, 1957.

RIGIDITY, INSECURITY, AND STRESS[1]

LEONARD H. AINSWORTH[2]
University College, London

THE controversy as to whether rigidity is a situational phenomenon or a generalized personality characteristic is the context from which the present investigation has arisen. Luchins (11, 12, 13, 14), whose water-jar Einstellung test has been used frequently as a measure of rigidity, has steadfastly emphasized situational factors as determinants of rigid performance in problem-solving situations. On the other hand, Rokeach (15, 16) and many others have sought to establish that rigid behavior in the problem-solving situation is attributable to a generalized personality characteristic that predisposes the individual to rigid attitudes and behavior in a broad variety of situations. It has been proposed that rigidity emerges as a defense against anxiety or as a result of defenses against anxiety. Some authors (e.g. 7, 8) imply that chronically anxious people tend to manifest generalized rigidity, while others (e.g. 18) view as primarily situational both the anxiety and the rigidity which serves to reduce it. Levitt (10), in reviewing the literature of the water-jar Einstellung test, finds so little consistency in the findings of many investigations that he concluded that the test is not a valid measure of any generalized trait of rigidity. Commenting on the influence of situational stress upon performance on the water-jar test, he says that "the over-all picture is beclouded."

The general hypothesis underlying this investigation is that rigidity is attributable to insecurity, and, more specifically, that problem-solving rigidity is a function of insecurity experienced in the problem-solving situation. Although on the surface this hypothesis appears to imply that rigidity is situational rather than personality-related, in fact insecurity is offered as a bridge-concept between the two extreme and contradictory views that have heretofore predominated.

The view of insecurity, stress, and rigidity that underlies the present investigation may be outlined as follows. Insecurity may be induced in an individual by the stressful aspects of a given situation, but the degree of insecurity he experiences is not directly proportional to the degree of stress as objectively estimated. First, the train of symbolic processes aroused in a previous insecurity-producing situation may prolong the experience of insecurity beyond the limits of the particular situation in which it was aroused, so that the individual enters a new situation already feeling insecure. Secondly, a situation that is nonstressful to many individuals may be stressful to others, either because they are relatively ill-equipped to deal with it, or because they perceive the situation as stressful through some similarity it bears to other situations in which insecurity was previously experienced.

Some individuals may be described as characteristically more insecure than others because they are insecure in a greater number and variety of situations. It seems reasonable that such an "insecure" person would be more likely than a relatively "secure" person to enter a given problem-solving situation in an insecure state, or to be made insecure by the demands of the task even under relatively nonstressful conditions. It is because insecurity may be viewed both as situational and as personality-related that it can serve a mediating function between the two apparently contradictory views of rigidity.

Rigidity is believed to arise as a defense against the insecurity experienced in a specific situation, but this does not imply that rigidity in this specific situation is related to a general ego-defensiveness, save through the intervening variable of insecurity. Thus it is not expected that a person with a "rigid personality structure," or one who makes massive use of defense mechanisms of various sorts, will be rigid in a problem-solving situation unless he experiences insecurity therein. To the extent that his general defenses prevent him from ex-

[1] This report is an abridgement of a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of London. The author wishes to express his appreciation to his advisor, Dr. Roger W. Russell, for his helpful criticisms in the design of the experiment and for his advice throughout the original study.

[2] Now with the Medical Service, Supreme Bench, Baltimore, and the Johns Hopkins University.

periencing insecurity in the specific situation, he may perform flexibly.

In order to explore the validity of this general position, the following specific hypotheses were formulated:

1. The more insecurity that the individual experiences in his general life adjustment, the more likely he is to behave rigidly in a problem-solving situation. The assumption here is that the individual may carry insecurity over from other situations to the problem-solving situation.

2. The experimental introduction of stress in the problem-solving situation will increase the likelihood that the individual will behave rigidly therein, and the greater the stress the greater is the likelihood that he will behave rigidly. The assumption here is that the greater the stress the more insecure the individual will feel in the specific situation.

3. Defense mechanisms developed in general life adjustment are not directly related to rigid behavior in the problem-solving situation. To the extent that they are related to rigidity it is through insecurity operating as specified in the first hypothesis.

4. The introduction of stress tends to act differentially, according to the extent to which insecurity is experienced by the individual in his general life adjustment. This hypothesis implies that the two independent variables of insecurity and stress act together in affecting the dependent variable of rigidity.

METHOD

Criteria of Rigidity

The criteria of rigidity employed in this study are based on persistence in employing an Einstellung formula in the problem-solving situation presented by a water-jar test similar to that designed by Luchins (11). The general features of the test have been described in the literature and will not be repeated here. This particular version of the test consists of twelve problems. Problems 1 to 6 are Einstellung problems that can be solved only by employing the relatively complex Einstellung formula (B − A − 2C). Problems 7 to 9 are known as critical problems that may be solved either by the Einstellung formula or in a simple and direct way. Problems 10 to 12 are extinction problems that can be solved simply and directly but not by the Einstellung formula.

Luchins (12, 13, 14) insists that the proper criterion of problem-solving rigidity is the use of the Einstellung formula in extinction problems for which it provides an incorrect solution. The critical problems do not require a shift of method, since the Einstellung formula provides a correct solution, and therefore, he argues, the person who uses the Einstellung formula in the solution of the critical problems is not necessarily rigid. Other workers have defined rigidity as the use of the Einstellung formula in the solution of the critical problems, and have measured rigidity in terms of the number of critical problems that were solved by means of this formula. Rokeach (15, 16) adopts this second definition of rigidity and does not include extinction problems in his version of the water-jar test.

In the present investigation the *S*s were classified according to their performance on the water-jar test in such a way that either of these two criteria of rigidity could be employed. *S*s were classified as either flexible (F) or as manifesting one of three degrees of rigidity (R_1, R_2, or R_3) according to the following specifications:

F. *S* used the Einstellung formula for the Einstellung problems only.

R_1. *S* carried the Einstellung formula over to the first critical problem, but thenceforward did not use it.

R_2. *S* used the Einstellung formula for two or three of the critical problems but for none of the extinction problems.

R_3. *S* used the Einstellung formula in one or more of the extinction problems, offering it as his final solution, with no indication that he recognized that it did not provide a correct solution.

With this classification it is possible to use either criterion of rigidity:

1. *S*s classified as R_3 are rigid, and all others are considered flexible.

2. *S*s classified as R_1, R_2, and R_3 are rigid, and the three groups represent three degrees of rigidity.

Each hypothesis is subjected to two tests, corresponding to the two criteria of rigidity that have been proposed.

Criteria of Stress

Four groups of *S*s were tested for problem-solving rigidity under different conditions. Group A, the nonstress group, was tested under relaxed conditions with no time limit, and the three stress groups, Groups B, C, and D, received respectively one, two, and three increments of stress. The conditions of stress were intended to be mild and to approximate the degree of stress found in typical university examinations.

First increment. After giving the standard instructions for the water-jar test, *E* emphasized the importance of speed of performance and announced a time limit of ten minutes. At the end of three minutes he called out "only five minutes to go," and at the end of six minutes he said "three minutes left to go." He then remained silent until all *S*s had completed the test, and made no comment about the discrepancy between the announced time limit and the time it actually took for all *S*s to finish. (The longest time taken by any *S* was 22 minutes and the shortest 6 minutes.) Although the main stress was intended to be time pressure, the announcements introduced distraction, and a slight competitive note may also be involved in the group situation.

Second increment. After the speed instructions, *E* urged the *S*s to try to do well since the results of the test had been found to be related to academic success. This was intended to increase stress by increasing the degree of ego-involvement. In addition, a competitive com-

ponent was introduced by instructing the *Ss* to indicate completion of the test by raising the hand.

Third increment. A confederate was introduced as an expert in mathematics, who was interested in observing the *S*'s solution of problems. *S* was then given the test instructions and the special instructions for the first and second increments of stress. *E* and his confederate then took up positions on either side of *S*, looking over his shoulder while he worked.

The increments of stress were thus applied additively, without implying equality of the degrees of stress involved in each of the increments. Groups A, B, and C were tested in a group situation. The *Ss* of Group D were tested individually.

Criterion of Insecurity

The criterion of insecurity was the total score derived from combining the scores on the insecurity subtests of four tests of security-insecurity constructed by Ainsworth and Ainsworth (2). In these tests, insecurity was given an entirely subjective definition in terms of feelings reported by *S* as characteristic of himself. The areas of adjustment covered by these four tests were familial relationships, extra-familial relationships, avocations, and philosophy of life. Each of the insecurity subtests included 12 items, making 48 insecurity items in all. A score of one was given for each of the items answered "true" and a score of ½ for each "?" answer. *Ss* were classified into three subdivisions on the basis of their total insecurity scores.

"*Secure.*" Scores of 0 to 8.

"*Middle.*" Scores of 8½ to 14½.

"*Insecure.*" Scores of 15 to 32.

Criterion of Defensiveness

Each of the tests of security-insecurity devised by Ainsworth and Ainsworth (2) included a subtest of 12 items designed to measure the extent to which defense mechanisms were used in an attempt to reduce feelings of insecurity. A combined score was derived from these four subtests in the same way as the combined insecurity score was derived; this was the criterion of defensiveness employed in the present investigation. *Ss* were classified into three subdivisions, as follows:

"*Low defensiveness.*" Scores of 0 to 12.

"*Moderate defensiveness.*" Scores of 12½ to 18.

"*High defensiveness.*" Scores of 18½ to 36.

Subjects

The research was undertaken in two phases. In Phase I, the tests of security-insecurity were administered to a total pool of 173 *Ss*, first- and second-year students attending the University of London. From this pool 30 matched quartets were selected, the *Ss* who composed each quartet belonging to the same insecurity subdivision and the same defensiveness subdivision. Thus, one quartet might be "secure" and of "low defensiveness," another "secure" and of "moderate defensiveness," and so on. One member of each quartet was then assigned to each of the four groups, A, B, C, and D. In Phase II of the experiment, the four groups were then tested for rigidity under four different sets of conditions, as described earlier.

As a check on matching procedures, group means were calculated, and the mean differences between groups were found for the measures of insecurity and defensiveness. All the mean differences were found to fall well below the .05 level of significance, and it was concluded that the groups were well matched. There were no significant differences between the groups in sex distribution, and although there were minor uncontrolled age differences, these were found to have no appreciable influence upon the results (1).

TABLE 1

INSECURITY AND RIGIDITY

Rigidity Subdivisions	Insecurity Subdivisions			Total
	Secure	Middle	Insecure	
F	9	5	1	15
R_1	7	7	4	18
R_2	12	18	6	36
R_3	8	18	25	51
Totals	36	48	36	120

Note.—1st test: $\chi^2 = 17.01$; $p < .01$; *df* 2 (F + R_1 + R_2 vs. R_3). 2nd test: $\chi^2 = 25.57$; $p < .01$; *df* 6.

RESULTS

Each hypothesis is subjected to two tests, corresponding to the two criteria of problem-solving rigidity that have been proposed. According to the first criterion, only *Ss* classified as R_3 are considered to be rigid, and the test involves the comparison of R_3 *Ss* with the rest of the sample. According to the second criterion, *Ss* classified as R_1, R_2, and R_3 are all considered to be rigid and to represent three degrees of rigidity. The test employing this criterion of rigidity involves the whole distribution of *Ss* as classified as F, R_1, R_2, or R_3.

Hypothesis 1: Insecurity and Rigidity

Table 1 shows the frequency of *Ss* classified as F, R_1, R_2, and R_3 in each of the "secure," "middle," and "insecure" subdivisions. The first test of the hypothesis involves the comparison of *Ss* classified as rigid (R_3) with the rest of the sample. It may be seen that this extreme form of rigidity tends to be characteristic of the insecure subdivision, and to occur less frequently in the middle subdivision, and still less frequently in the secure subdivision. The null hypothesis is rejected at the .01 level of significance.

The second test of the hypothesis involves the whole distribution. It may be seen that flexibility (F) and the least degree of rigidity (R_1) are relatively more frequent in the secure

TABLE 2
STRESS AND RIGIDITY
(*N* for each group is 30)

Rigidity Subdivisions	Control Group A	Stress Groups			Total
		Group B	Group C	Group D	
F	6	4	4	1	15
R_1	5	5	4	4	18
R_2	14	8	6	8	36
R_3	5	13	16	17	51

Note.—1st test: $\chi^2 = 10.12$; $p < .02$; df 3 (F + R_1 + R_2 vs. R_3). 2nd test: $\chi^2 = 14.31$; $p < .10$; df 9.

TABLE 3
DEFENSE MECHANISMS AND RIGIDITY

Rigidity Subdivisions	Defensiveness Subdivisions			Total
	Low	Moderate	High	
F	6	7	2	15
R_1	7	6	5	18
R_2	12	15	9	36
R_3	15	14	22	51
Totals	40	42	38	120

Note.—1st test: $\chi^2 = 5.52$; $p < .10$; df 2 (F + R_1 + R_2 vs. R_3). 2nd test: $\chi^2 = 6.99$; $p < .50$; df 6.

than in the middle subdivision, and more frequent in the middle than in the insecure subdivision. A moderate degree of rigidity (R_2) is more frequent in the middle subdivision than in either the secure or insecure subdivisions. And, as mentioned above, the most extreme form of rigidity (R_3) occurs most frequently in the insecure subdivision. The departure of this distribution from one that might be expected on the basis of chance is significant at the .01 level of significance.

The hypothesis is thus supported that rigidity in a problem-solving situation is related to insecurity in general life adjustment.

Hypothesis 2: Stress and Rigidity

Table 2 shows the frequency of *S*s classified as F, R_1, R_2, and R_3 in the nonstress group, Group A, and in each of the three stress groups, Groups B, C, and D. The first test of the hypothesis involves the comparison of *S*s classified as rigid (R_3) with the rest of the sample. It may be seen that rigidity thus defined is more frequent in the stress groups than in the nonstress group, and that rigidity increases in frequency with increasing stress. This tendency is significant at the .02 level of significance.

It is obvious that the greatest differences lie between the nonstress group, Group A, and the three stress groups, whereas the differences between the stress groups themselves are relatively small. Chi squares were calculated for each pair of groups separately, and Group A was found to differ significantly from each of the other groups with respect to R_3 behavior, but there were no other significant differences.

The second test, involving the whole distribution, does not support the hypothesis. Although there is a tendency for flexibility (F) and the milder forms of rigidity (R_1 and R_2) to decrease with increasing stress, as well as for the more extreme form of rigidity (R_3) to increase with increasing stress, the probability that this departs from a chance distribution is between .05 and .10, so that the null hypothesis cannot be rejected with confidence.

As a check on this finding, chi squares were calculated for each pair of groups separately, and the only significant departure from chance expectations was found when Group A was compared with Group D, the group receiving most stress.

Thus, only partial support for the second hypothesis is found. The experimental introduction of stress in the problem-solving situation increases the likelihood that the individual will behave rigidly therein, if rigidity is defined as the failure to shift when the problem requires a shift. The hypothesis is not supported if the definition of rigidity includes the milder forms in which a shift is not made when the problem permits but does not require a shift. Support is not found for that part of the hypothesis which states that rigidity will increase with increasing stress.

Hypothesis 3: Defensiveness and Rigidity

Table 3 shows the frequency of *S*s classified as F, R_1, R_2, and R_3 in each of the subdivisions of defensiveness—low, moderate, and high. The hypothesis that rigidity is not related to degree of defensiveness is supported by the first test, in which the *S*s classified as R_3 are compared with the rest of the sample. Although there is a tendency for R_3 behavior to occur more frequently among *S*s with high defensiveness than among those with moderate or low defensiveness, this tendency falls short of statistical significance.

When the hypothesis is submitted to the

second test, involving the whole distribution, it is clear that the null hypothesis cannot be rejected. According to both tests, the hypothesis is thus supported that rigidity in a problem-solving situation is unrelated to degree of defensiveness in general life adjustment.

Hypothesis 4: Insecurity, Stress, and Rigidity

In Fig. 1, the percentages of *S*s classified as F, R_1, R_2, and R_3 are shown for each of the "secure," "middle," and "insecure" subdivisions of the nonstress group and each of the three stress groups. Percentages are used instead of raw frequencies because there were more *S*s in the middle subdivisions than in the secure and insecure subdivisions.

In order to demonstrate that the independent variables of insecurity and stress act together in determining the incidence of rigidity it is necessary that (*a*) in each of the stress groups, Groups B, C, and D, there be a significant increase in the incidence of rigidity from the secure to the middle and from the middle to the insecure subdivision, and that (*b*) in each of the secure, middle, and insecure subdivisions there be a significant increase in the incidence of rigidity from Group B to Group C, and from Group C to Group D.

In the first test of the hypothesis the relative frequency of R_3 behavior in the various subdivisions of the four groups of *S*s is of concern. In Group A, the nonstress group, the only *S*s who are classed as R_3 are insecure *S*s. In each of the three stress groups, *S*s classed as R_3 occur in the secure and middle subdivisions as well as in the insecure subdivisions. In each of these stress groups more middle than secure *S*s and more insecure than middle *S*s are classed as R_3. In each case the findings are thus exactly in accordance with expectations. Only once in 216 times, i.e., $(1/6)^3$ could this predicted order of subdivisions, (secure < middle < insecure), be expected to occur by chance in three independent samples. Thus the first half of the hypothesis is supported at a highly significant level; the more insecure the person, the more likely he will be to behave rigidly under stress. This effect is consistent at all three levels of stress represented in this study.

[3] These three *S*s were discarded from the sample and their places taken in their respective quartets by three other *S*s from the pool.

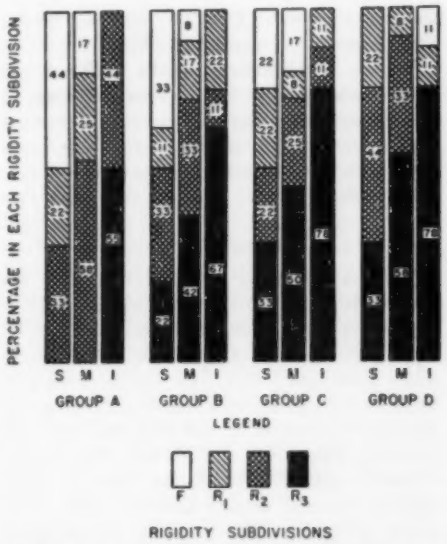

FIG. 1. PERCENTAGES OF THE SECURE, MIDDLE, AND INSECURE SUBDIVISIONS OF EACH OF THE FOUR STRESS GROUPS SHOWING EACH OF FOUR DEGREES OF RIGIDITY

The second half of the interaction hypothesis provides that the greater the stress, the greater the incidence of rigidity. The hypothesis holds good when Group B is compared with Group C. In the case of each of the secure, middle, and insecure subdivisions, the incidence of rigidity is greater in Group C than in Group B. But when Group C is compared with Group D, expectations are met only in the case of the middle subdivision, in which Group D is more rigid than Group C. The incidence of rigidity in Group C and Group D is equal for both of the secure and insecure subdivisions. Therefore, the second half of the hypothesis is not supported, since there is no significant tendency for incidence of rigidity to increase with increased stress.

In the second test of the hypothesis, all degrees of rigidity (R_1, R_2, and R_3) are involved, in comparison with each other and with flexibility (F). It may be seen that in both Group B and Group C the secure subdivision is more flexible than the middle subdivision, and the middle subdivision is more flexible than the insecure subdivision. This is the case both when one considers the *S*s classified as F alone, and when F and R_1 are combined. But the same tendency does not

hold for Group D. Thus the first half of the hypothesis is not supported. When F and R_1 incidence is examined within each of the secure, middle, and insecure subdivisions, it may be seen that there is no regular tendency for flexibility to decrease with increasing stress. Thus the second test fails to support the hypothesis.

DISCUSSION

This investigation demonstrates that the extent to which situational influences affect rigidity depends upon the criterion of rigidity employed. The first criterion used in this study follows Luchins (11, 12, 13, 14) in classing as rigid only those *S*s who fail to shift to a new method of solution of the problem when the situation demands a shift. When this criterion is used, support is found for Luchins' claim that situational factors influence the occurrence of rigidity. The introduction of experimental stress is found significantly to increase the failure to shift when the situation demands a shift.

The second criterion distinguishes degrees of rigidity in inverse proportion to the readiness with which a shift to a new method of solution is made when the situation permits but does not demand a shift. This criterion approximates that used by Rokeach (15, 16). When the milder degrees of rigidity are taken into consideration, the influence of situational factors upon rigidity are obscured. The introduction of experimental stress does not significantly decrease the readiness to shift before the situation demands a shift.

It is suggested that some of the apparent contradictions in the literature of the water-jar test may be attributable to differences in the criteria of rigidity employed. Until a consensus is reached regarding the basic conditions that influence the incidence of problem-solving rigidity, it would seem wise to adhere to the more extreme criterion, namely, failure to shift when the situation demands a shift. It is clear that this behavior may be described as rigid. It is not clear that it is rigid to use a relatively complicated solution instead of a relatively direct solution to a problem, when both solutions are correct, and when the more complicated solution may well be the more economical because the preceding problems have made it the solution most readily available to the person.

When rigidity is defined as failure to shift when the situation demands a shift, it is found to be influenced both by situational factors and by the degree of insecurity experienced by the person in his general life adjustment. Indeed, under the conditions of the present experiment, insecurity characteristic of the person seems to play a larger part in determining rigid behavior than does insecurity induced by situational manipulation.

Under the conditions of the present experiment, the incidence of rigidity is not found to increase significantly with an increase in stress, although there are significant differences in incidence of rigidity in stress situations in contrast with the nonstress situation. It seems probable that a closer correspondence between increasing stress and increasing incidence of rigidity could be demonstrated in an experiment in which the stress increments were larger. It seems a reasonable hypothesis that the introduction of severe stress might override or cloak the contribution made by personality-related insecurity. This hypothesis implies that there is a limit beyond which even the relatively secure individual can not maintain flexibility under stress. In the present investigation this limit is not approached, since only a third of the secure *S*s behave rigidly under the most severe stress applied here.

It is relevant to the discussion of limits of stress-tolerance to draw attention to the disorganization that occurs with a few Ss, a finding to which reference has not heretofore been made. In some of the problems of the water-jar test, ten *S*s neither used the Einstellung formula nor the simple, direct solution; instead, they attempted long and complicated solutions that were arithmetically incorrect, and further betrayed disorganization by other signs such as scratching out false starts. Seven of the ten were classified as R_3, and nine were in the stress groups. They were distributed irregularly through the security-insecurity subdivisions. Furthermore, there were three other *S*s who were discarded[4] because their water-jar test performances were unscorable

[4] See Footnote 3.

according to the criteria used. Two of these three *S*s were clearly disorganized. These were tested under Group D conditions where their behavior could be observed. They trembled and seemed completely blocked and close to tears. (The third *S* may also have been disorganized to the point of complete blocking for he attempted none of the problems; this *S* was tested under group conditions, so it was not possible to ascertain the reason for his failure.)

The phenomenon of disorganization has not heretofore been reported in the literature of the water-jar test. This seems strange, since the literature on stress reports, as the most immediate and striking effect of stress, a disorganized and diffuse pattern of response (e.g., 3, 9, 17). It is suggested that two practices with respect to the water-jar test have tended to obscure the emergence of disorganization: (*a*) the general practice of administering the test in a group situation in which the behavior of the *S* can be observed only through his pencil-paper record, and (*b*) the general practice of discarding from the data test records that include arithmetical errors other than those attributable to the Einstellung effect, thus discarding the evidence for disorganization.

It is suggested that, to the extent that the individual experiences insecurity in the problem-solving situation, this insecurity manifests itself in tendencies towards a diffuse and disorganized response, even though these tendencies are not easily observable in the test record under group test conditions. The Einstellung formula offers a relatively secure alternative to disorganization, and hence the disorganized *S*s tend to return to it, even though in so doing they avoid tackling the solution to the problem. In this way, rigidity may be viewed as a defense against insecurity experienced in the problem-solving situation. Thus it is believed that rigidity may emerge as a defense in a specific situation even though it is not found to be related significantly to the degree of defensiveness in general life adjustment.

It is suggested that the positive relationship between rigidity and general defensiveness reported frequently, but by no means unanimously, in the literature may be accounted for by the fact that highly defensive persons tend to be relatively insecure. This was clearly demonstrated in the case of the measures used in the present study by Ainsworth and Ainsworth (2) who found significant positive relationships between scores on all the insecurity subtests of the security-insecurity tests and the subtests measuring degree of defensiveness. The hypothesis consistent with this writer's position is that to the extent that defense mechanisms are effective in protecting the individual against feelings of insecurity, he should be able to meet a relatively nonstressful problem-solving situation without appreciable insecurity and deal flexibly with it. On the other hand, if the person is relatively insecure despite considerable use of defense mechanisms—and this is more usually the case—he would be rigid in a problem-solving situation. These hypotheses are supported in the present investigation, since rigidity is found to be significantly related to insecurity but not to degree of defensiveness.

SUMMARY

1. The general hypothesis of the study is that problem-solving rigidity is attributable to insecurity experienced in the problem-solving situation. This hypothesis is believed to offer a resolution of the controversy as to whether rigidity is a situational phenomenon or attributable to a generalized personality characteristic. Insecurity is the bridge-concept between these two apparently contradictory views, since the degree of insecurity experienced in a specific situation is held to be influenced both by the stressfulness of the situation itself and by the degree of insecurity experienced by the individual in his general life adjustment.

2. Four specific hypotheses were formulated and may be summarized as follows: (*a*) rigidity is related to insecurity in general life adjustment; (*b*) rigidity is related to stress; (*c*) rigidity is not related to degree of defensiveness in general life adjustment; and (*d*) the introduction of stress tends to act differentially in producing rigidity in accordance with the degree of insecurity experienced in general life adjustment.

3. *S*s were 120 university students divided into four groups of 30 each, matched on the

basis of insecurity and defensiveness in general life adjustment. A water-jar test of problem-solving rigidity was administered to each group under different conditions, one group serving as a control group and the other three groups receiving respectively one, two, and three increments of stress. Two alternative criteria of rigidity were employed. The first criterion defined rigidity as a failure to shift to a new method of solution when the problem requires a shift. The second criterion distinguished two milder degrees of ridigity as well as the more extreme form of rigidity implied by the first criterion.

4. Significant support is found for the hypothesis that rigidity is related to insecurity in general life adjustment. Support is also found for the hypothesis that rigidity is related to situational stress, but only when rigidity is defined as a failure to shift when the problem requires a shift, and not when milder degrees of rigidity are also considered. Although the incidence of rigidity is significantly greater in the stress groups than in the nonstress groups, the incidence does not increase significantly as increments of stress are added. Because of this latter finding it is not possible to demonstrate interaction between the two variables of stress and insecurity, but the relationship between rigidity and degree of insecurity in general life adjustment is found to hold under three different degrees of stress. Support is found for the hypothesis that rigidity is not significantly related to degree of defensiveness in general life adjustment.

5. The implications of these findings are discussed, with particular stress upon the influence of the criterion of rigidity employed. Attention is drawn to the phenomenon of disorganization that occurred with a few *S*s, and the hypothesis is formulated that insecurity in the problem-solving situation manifests itself in tendencies towards a diffuse and disorganized response, and that it is against these tendencies that rigidity acts as a defense. This hypothesis does not imply a generalized rigidity factor characteristic of the personality.

REFERENCES

1. AINSWORTH, L. H. A study of rigidity. Unpublished doctoral thesis, University College, Univer. of London, 1953.
2. AINSWORTH, MARY D., & AINSWORTH, L. H. *Measuring security and insecurity.* Toronto: Univer. of Toronto Press, in press.
3. BARKER, R. G., DEMBO, T., & LEWIN, K. Frustration and regression: An experiment with young children. *Univer. Ia. Stud. in Child Welf.*, 1941, **18**, No. 1.
4. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
5. CATTELL, R. B. The riddle of perseveration. II. Solution in terms of personality structure. *J. Pers.* 1946, **14**, 239–267.
6. CATTELL, R. B., & TINER, G. The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
7. ERIKSEN, S. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1952–53, **21**, 386–391.
8. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
9. GOLDSTEIN, K. *The organism.* New York: American Book, 1939.
10. LEVITT, E. E. The water-jar Einstellung test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.
11. LUCHINS, A. S. Mechanization in problem-solving. *Psychol. Monogr.*, 1942, **54**, No. 6.
12. LUCHINS, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17**, 449–466.
13. LUCHINS, A. S. On recent usage of the Einstellung effect as a test of rigidity. *J. consult. Psychol.*, 1951, **15**, 89–94.
14. LUCHINS, A. S. The Einstellung test of rigidity: its relation to concreteness of thinking. *J. consult. Psychol.*, 1951, **15**, 303–310.
15. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
16. ROKEACH, M. Prejudice, concreteness of thinking and reification of thinking. *J. abnorm. soc. Psychol.*, 1951, **46**, 83–91.
17. SELYE, H. *The physiology and pathology of exposure to stress.* Montreal: Acta, 1950.
18. WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1954, **48**, 129–134.

Received April 12, 1957.

INTELLIGENCE AND EMOTIONAL DISTURBANCE: EVIDENCE FROM ARMY AND VETERANS ADMINISTRATION RECORDS[1]

JOHN B. MINER[2] AND JAMES K. ANDERSON

Conservation of Human Resources Project, Columbia University

FOR many years, psychologists have been attempting to determine whether or not there is a difference on the average in intellectual level between individuals suffering from emotional disturbance and the normal population. The first studies appeared during World War I, and interest continued into the 1920's. Then, after a hiatus in the late twenties and early thirties, a flood of publications on the subject appeared as large numbers of psychologists began to accept positions on mental hospital staffs. By 1945, however, this activity had virtually ceased. Since World War II only a handful of studies has appeared, most of which have been tangential by-products of more broadly conceived investigations.

The decline in research activity seems to have resulted not from a lack of interest, but from an increased understanding of certain basic methodological difficulties. The results of the earlier studies were far from conclusive. In some instances, lower intelligence was found throughout all the emotionally disturbed groups studied. Other investigations reported lower test scores only among certain diagnostic groups. Still others failed to establish any differences, and, on occasion, pathological groups were well above the general level of the population in intelligence. It became apparent that a definitive answer cannot be obtained by testing a group of patients in a single hospital and comparing their scores with a control group whose representativeness is equally questionable.

However, sample selection was not the only factor contributing to the inconclusive results. As Roe and Shakow (8) demonstrated, the test scores of psychotics and even neurotics are frequently lowered by a variety of distractors that are much more closely associated with the psychiatric condition itself than with the level of intellectual functioning. Such distractors include those extraneous thoughts, feelings, and perceptions that intrude upon consciousness and thus interfere with a total concentration on the test items. Thus, hallucinations, obsessive ruminations, anxiety attacks, and flights into fantasy frequently occur in the test situation. Accordingly, performance is often spotty with easy items failed and difficult ones passed. The scores on timed tests are particularly likely to be lowered by these distractors. Certainly estimates of intellectual level obtained under such conditions cannot be considered as valid indicators of optimum or even typical functioning. Furthermore, if patients must be tested during a period of emotional illness, as when the problem of intellectual deterioration is under investigation, controls must be provided that will permit a clear differentiation between the effects of distractors and any possible deterioration effects.

Although the results of previous studies fail to establish a lower level of intelligence as characteristic of the emotionally disturbed, certain facts suggest that this is a tenable hypothesis. Since 1949, the staff of the Conservation of Human Resources Project has been studying military manpower utilization during World War II. This work has already resulted in the publication of two books, *The Uneducated* (4) and *Psychiatry and Military Manpower Policy* (5). Three additional volumes dealing with ineffective performance as related to the military and post-war civilian environments are now in preparation. In the course of these investigations it has become evident that those servicemen who were discharged for psychosis and neurosis were as a group well below the army as a whole in educational attainment at the time of induction. Comparative data are presented in Table 1. It is also clear that there is a high positive relationship between education and intelligence. This conclusion has been substantiated elsewhere (7) employing evidence derived from a representative sample of

[1] Drs. Eli Ginzberg Director of the project, Sol W. Ginsburg consultant in psychiatry, and John L. Herma staff psychologist reviewed this paper and made valuable suggestions. Similar contributions were made by Drs. Irving Lorge of Teachers College, Columbia University, Silvan S. Tomkins of Princeton University, and Douglas W. Bray of the American Telephone and Telegraph Company.

[2] Now with the Industrial Relations Department, Atlantic Refining Co., Philadelphia, Pa.

TABLE 1

PER CENT OF ARMY NEUROTICS, PSYCHOTICS, AND TOTAL INDUCTEES ATTAINING VARIOUS EDUCATIONAL LEVELS

Group	School Years Completed			
	0–8	9–11	12	13+
Neurotics	41	31	20	8
Psychotics	40	30	20	10
Inductees	32	28	26	14

the United States population. Other investigators have obtained correlation coefficients varying generally in the range .50 to .75 using more restricted samples. The median value appears to be approximately .65. These findings suggest the following hypothesis:

1. The intelligence of neurotics and psychotics prior to the time of breakdown is below that of the population as a whole.

In the belief that intelligence may in many cases be an important factor in the resolution of emotional conflicts at the conscious level and thus in a recovery, a second hypothesis was formulated.

2. The pre-breakdown intelligence of individuals who later develop a severe psychiatric disorder of a *chronic* nature is below that of individuals who do not *remain* severely disturbed.

METHOD

In the investigation of these hypotheses, materials collected in connection with the Conservation of Human Resources Project's study of ineffective performance during World War II were employed. The cases derive from a special sample of World War II enlisted men retained by the Adjutant General's Office in Washington. From this sample, which consisted of five per cent of all enlisted separations from the army, all individuals inducted during the last four months of 1942 who had been prematurely discharged for inaptness, undesirability, psychosis, and psychoneurosis were selected. From this prematurely discharged group, special samples were selected to be representative within the following categories:

Service in the U. S. of less than six months.
Service in the U. S. of six to 18 months.
Service in the U. S. of 18 months or more.
Service overseas in the ground forces.
Service overseas in the Air Force.
Service overseas in the service forces.

There were 231 psychoneurotics and 120 psychotics included. By weighting the data within categories, it was possible to obtain figures representative of enlisted men inducted during the last four months of 1942. Approximately one sixth of the over ten million army accessions in World War II occurred during this period.

Extensive information on the subjects (*Ss*) was obtained from the files of the Army Records Center in St. Louis. These files contain service records, medical records, and separation board proceedings. Follow-up information has also been made available from the files of the Veterans Administration. Many of the *Ss* obtained disability compensation and medical treatment from the Veterans Administration. Others received extensive education and training, frequently coupled with vocational guidance interviews and testing. These contacts have resulted in the accumulation of a variety of records that in most cases give a comprehensive picture of the veterans' adjustment since discharge. When the Veterans Administration data were, for some reason, incomplete, they were supplemented through the use of questionnaires mailed to the veterans' homes. In general, however, since nearly all soldiers discharged for psychosis and many psychoneurotics received disability compensation for a number of years and were accordingly examined at frequent intervals, the VA records on these groups are quite complete as checked through 1953.

Because many of the Form 20s, which listed Army General Classification Test scores for each soldier, were destroyed after the War, it was impossible to obtain information on the intelligence of all *Ss*. Scores for tests administered at the time of induction were available on 103 (44.6 per cent) of the neurotics and 61 (50.8 per cent) of the psychotics. These samples could have been supplemented if intelligence test scores listed in the medical records and obtained during a period of hospitalization for emotional disturbance had been included. However, this would have contaminated the results by introducing an unknown number of scores that had been artificially affected by disease-related distractors. Although most of the AGCT results were obtained from army records, there were some instances in which they were available only in the VA files. It is, of course, well known that testing conditions at the various induction centers were not completely standardized. Nevertheless, the error thus introduced does not appear to be sufficiently large or sufficiently systematic to invalidate any major conclusions.

The diagnostic composition of the sample is indicated in Table 2. These diagnoses were given at the time of discharge from service and were obtained from the medical records. Because of the lack of any standard method of classification that could be applied throughout World War II and the very sizable differences in training and experience among the psychiatrists making the diagnoses, these data can be considered as suggestive only. The major division into psychoneurotic and psychotic groups is, however, in general quite consistent with the impressions derived from a thorough reading of the medical histories by several Conservation of Human Resources Project staff members and appears to be reliable. It should be emphasized that these samples contain only those men who were discharged from the Army and who received a psychiatric diagnosis. This diagnosis in most cases reflected a severe disturbance in the ability to perform. However, there were some men who, although they did suffer from symptoms of an emotional origin, were discharged only because their symptoms came to the attention of the

TABLE 2
DISCHARGE DIAGNOSES IN NEUROTIC AND PSYCHOTIC SAMPLES

Diagnosis	*N*	%
Psychoneurosis		
Anxiety state	54	52.4
Mixed type	13	12.6
Conversion hysteria	11	10.7
Hysteria	7	6.8
Neurocirculatory asthenia	5	4.9
Neurasthenia	4	3.9
Reactive depression	4	3.9
Obsessive-compulsive	2	1.9
Unspecified type	2	1.9
Hypochondriasis	1	1.0
Psychosis		
Dementia praecox, paranoid	14	23.0
Dementia praecox, unspecified	11	18.0
Unspecified type	8	13.1
Dementia praecox, simple	6	9.8
Dementia praecox, catatonic	4	6.6
With Mental Deficiency	4	6.6
Schizo-affective type	3	4.9
Dementia praecox, hebephrenic	3	4.9
Dementia praecox, mixed	2	3.3
Manic-depressive, manic	2	3.3
With Psychopathic Personality	2	3.3
Manic-depressive, depressed	1	1.6
With Epilepsy	1	1.6

authorities during a period when the army's personnel policy made separation relatively easy. These individuals, whose performance was not grossly affected by their symptoms, were more likely to be discharged for psychoneurosis than for psychosis.

Information on the AGCT scores for total inductions in the period September, 1942, through May, 1943, was obtained from Davenport (3). Because the neurotic and psychotic samples contained only enlisted men these data were corrected to exclude men who later became army officers. This control group differs from the experimental samples in that it includes men inducted during the first five months of 1943. There is no reason, however, to believe that the early 1943 inductees differed in AGCT score from those who entered the army during late 1942. It was not possible to eliminate from the control group those men who were subsequently discharged for psychiatric reasons.

In order to investigate the first hypothesis, comparisons were made between the distributions of AGCT scores for the neurotics, psychotics, and controls. The distributions for the two pathological groups were obtained by weighting the data within each of the six length-of-service and branch-of-service categories according to the incidence of neurosis and psychosis in each category. The weighting factors were derived from the five per cent sample retained by the Adjutant General's Office. Significance was tested by the chi square method because control data were available only in terms of gross score ranges.

As previously indicated, any definitive test of this hypothesis must employ an experimental design that rules out the effects of biased sampling and disease-related distractors as well as any possible deterioration effects. In the use of military data, a major sampling difficulty arises because AGCT scores are not available for all *S*s. As a check on the possibility of sample selection, comparisons on education, race, and age were made between the subsample whose AGCT scores were available and a group whose scores were lost. Since all of these variables have a sizable and, to some degree, independent relationship with intelligence (7), failure to establish differences between the two groups would seem to offer strong evidence in support of the hypothesis that selection on intelligence has not occurred. In order to investigate the possibility that emotional disturbance existing at induction may have lowered test scores, the AGCT distribution of those who broke down during the first three months of service was compared with the distribution for those whose illness appeared later. Breakdown was defined as any contact with the medical service, whether or not hospitalization was required, that resulted in a psychiatric diagnosis. The assumption was that if those who suffered from severe mental disorders were not diagnosed and rejected at the time of the induction examination, their condition would, nevertheless, in most instances become evident within the first three months of service. Failure to establish a difference in AGCT score between the early and late breakdown groups would indicate either that illness at induction did not have an effect on the intelligence test results of this group or that even those who broke down early were free of disease at the time of testing. It is, of course, true that failure to establish a difference might result if the early breakdown group had a higher pre-illness intelligence than the late breakdown group but was disturbed at the time of induction testing and thus reflected lowered scores. Such a situation seems highly improbable. In any event, a finding of no difference in intelligence between the early and late breakdown groups would seem to offer strong presumptive evidence in support of the conclusion that any differences found between the control and pathological groups could not be attributed to the effects of disease-related distractors or deterioration.

Finally, in order to investigate the second hypothesis, the AGCT distribution of those with a post-war history of hospitalization for emotional disorder was compared with the distribution for those with no such history. Postwar is understood to include the period from discharge through 1953. The data on hospitalization came almost exclusively from the VA records, although the questionnaires were also inspected. With the exception of three cases, all members of the hospitalized sample were at one time or another in VA facilities. In some cases, institutionalization followed immediately on discharge, whereas in others, a period of freedom from hospital care intervened. Multiple admissions were not uncommon. Since a number of men discharged from the service for psychoneurosis were later hospitalized for psychosis and a few discharged for psychosis later received a diagnosis of psychoneurosis, no attempt was made to differentiate between these two groups. Of the 64 *S*s hospitalized after the war, 16 were considered psychoneurotic at the time of separation from service and 48 psychotic. In view of the VA's continual interest in the psychiatric status of these men and the fact that the men themselves stood to benefit through increased pension payments by keeping the VA informed of any hospitalization for emotional disturbance, there is

TABLE 3
PER CENT OF NEUROTICS, PSYCHOTICS, AND CONTROLS SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Neurotics[a]	103	15.8	42.5	22.0	19.7
Psychotics[a]	61	22.2	33.2	32.4	12.2
Controls		10.6	31.8	33.0	24.6

[a] Based on weighted data.

little reason to believe the data on institutionalization are significantly in error.

RESULTS

The weighted percentage distribution for the neurotics and the psychotics and the percentage distribution for the controls are presented in Table 3. The standard grouping of AGCT scores was used except that Groups I and II (scores of over 129 and 110 through 129 respectively) were combined because of the small number of cases in each of these two categories among the psychiatric samples. The control *S*s differ significantly from both the neurotics and the psychotics. (The neurotic-control comparison yields a χ^2 of 11.16 with $p < .02$, while the psychotic-control comparison yields a χ^2 of 11.59 with $p < .01$.) It is clear that, in accordance with the first hypothesis, both psychiatric groups score well below the control sample. In addition, the neurotics and psychotics do not differ reliably from one another with respect to AGCT score (χ^2 of 4.55, $p > .10$). The mean weighted AGCT score for the neurotics is 81.2 and for the psychotics 84.7. Because control data are available only in terms of gross score ranges and because the control sample includes men who were later discharged for psychiatric reasons, it is difficult to specify a precise figure for the normal group. However, the mean is almost certainly not below 94 and may well be several points higher.

Data on the educational level attained by the *S*s whose AGCT scores were available and those for whom such scores were not available are presented in Table 4. When both psychiatric groups are combined, it is clear that those whose AGCT scores are available have a higher educational level (χ^2 of 8.46, $p < .05$). Within both the neurotic and psychotic groups the educational difference between those with and those without AGCT scores is in the same direction but does not attain significance (p values between .10 and .20 in both cases). Since, in addition, educational level and AGCT score have a high relationship with each other in both groups ($r = .74$ for the neurotics and .60 for the psychotics), it seems probable that the mean scores for the psychiatric groups are, if anything, too high. The educational selection of the particular samples used in this study appears to have resulted in an underestimate of the intelligence differential between those suffering from emotional disturbance and normals. It is possible that a similar selection factor was operative in other studies employing World War II data (2, 6) which yielded results somewhat at variance with those obtained here. However, without cross-validation the existence of a stable selective bias cannot be definitely established.

TABLE 4
PER CENT OF AGCT AND NON-AGCT SAMPLES ATTAINING VARIOUS EDUCATIONAL LEVELS

Sample	N	School Years Completed			
		0–6	7–8	9–11	12+
AGCT	164	17.7	18.3	31.1	32.9
Non-AGCT	187	26.2	24.6	26.7	22.5

TABLE 5
PER CENT OF AGCT AND NON-AGCT SAMPLES OF WHITE AND NEGRO RACE

Sample	N	Race	
		White	Negro
AGCT	164	81.7	18.3
Non-AGCT	187	85.0	15.0

As indicated in Table 5, the AGCT and non-AGCT samples differ little in racial composition (χ^2 for the total psychiatric group is .48, for the neurotics .01, and for the psychotics 1.68, all with $p > .10$). Similarly, the two samples do not differ with respect to age. (For the total psychiatric group shown in Table 6, the χ^2 is 3.46, for the neurotics alone 2.60, and for the psychotics alone 5.04, all with $p > .10$.)

Table 7 presents the AGCT distributions for those men who exhibited psychiatric symptoms during their first three months of service and for those who did not exhibit symptoms until later. The two groups do not differ significantly (χ^2 is .07, $p > .10$). The unweighted

TABLE 6
PER CENT OF AGCT AND NON-AGCT SAMPLES OF VARIOUS AGES AT INDUCTION

Sample	N	Years			
		18–24	25–29	30–34	35+
AGCT	164	64.0	14.0	12.2	9.8
Non-AGCT	187	65.2	19.3	8.6	6.9

TABLE 7
PER CENT OF EARLY AND LATE BREAKDOWN GROUPS SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Early breakdown	20	15.0	35.0	30.0	20.0
Late breakdown	144	14.0	33.3	32.6	20.1

TABLE 8
PER CENT OF THOSE HOSPITALIZED AND NOT HOSPITALIZED AFTER DISCHARGE SCORING AT VARIOUS AGCT LEVELS

Group	N	AGCT Scores			
		−59	60–89	90–109	110+
Hospitalized	64	26.6	35.9	28.1	9.4
Not hospitalized	100	6.0	32.0	35.0	27.0

mean for the early breakdown group is 87.5; for the later it is 88.4. These results strongly suggest that neither intellectual deterioration nor illness-related distractors affected test performance in such a way as to produce the differential noted between the AGCT scores of the psychiatric groups and the controls.

Evidence on the intelligence of those who were hospitalized for psychiatric disorders after separation from the Army as compared with the nonhospitalized group is presented in Table 8. There is a clear-cut tendency to lower scores among the hospitalized group (χ^2 is 17.38, $p < .01$). The mean AGCT score of the hospitalized group is 78.3 and of the nonhospitalized 95.8. This result supports the second hypothesis.

DISCUSSION

Strictly speaking, these findings can only be generalized to that portion of the World War II army that was inducted during the last four months of 1942. Since, however, this includes approximately one sixth of those who served, and since there is no reason to believe that this group differed in any respect relevant to the results from the army as a whole, the conclusions can, with some degree of confidence, be applied to the total group who served in World War II. It seems probable that similar results would be obtained in a study based on the population as a whole. However, due to the restricted age range of the military group, as well as its higher educational level, this generalization cannot be made with certainty. Furthermore, it cannot be assumed without further study that identical results would be obtained if a different set of operations for defining intelligence, e.g., a test other than the AGCT, was employed.

Although the fact of a lower level of intelligence among the emotionally disturbed groups has been established, the findings offer little in the way of clues to an understanding of why this is the case. Theoretically, there are three possibilities. First, limited intelligence may, under certain circumstances, contribute to the development of a psychiatric disorder. Second, character disturbances existing at an early age and later eventuating in a true emotional disorder may act to inhibit learning and the development of native potential. Third, some other factor or factors may operate to produce both a lowered level of intelligence and a predisposition to breakdown.

Although further study is certainly required, it seems probable that all three factors are operative. The case materials collected in connection with the Conservation of Human Resources Project studies of ineffective performance are particularly revealing with regard to the hypothesis that limited intelligence may contribute to the development of a psychiatric disorder. Although many men were unable to perform adequately in the military situation because of marked vulnerability under stress (low ego strength or emotional instability), and many others were severely lacking in the necessary intelligence, there was, in addition, a third group in which these two factors seemed to overlap. These men apparently failed because in the military setting they were exposed for the first time to demands for learning and understanding beyond their intellectual competence. They became threatened, anxious, and eventually their performance was severely affected. It seems likely that the lower the level of a man's intelligence, the greater is the probability that, at some time in his life, he will be

exposed to environmental demands with which he is unable to cope. If these demands become so insistent as to make him acutely aware of his intellectual deficiencies, intense feelings of inadequacy may be precipitated and a severe emotional disturbance ensue.

It also seems probable that some types of character disturbances which are of insufficient severity to be considered true emotional disorders may operate from an early age to inhibit the development of innate learning ability. Evidence in support of this conclusion has been presented by Weisskopf (10) and by Sontag, Baker, and Nelson (9). Motivation to learn may not develop or may be misdirected so that the knowledge that our society values and thus includes in the school curriculum and intelligence tests is not absorbed. Since, in such individuals, ego strength may be assumed to be minimal and susceptibility to stress high, the probability of emotional breakdown is well above that of the population as a whole. Thus, early character disturbances associated with low ego strength would seem not only to predispose to later severe maladjustment but also to limit the development of intellectual capacities.

The third notion predicates such factors as a generally deprived environment as producing a lower intelligence among the emotionally ill. Elsewhere (7) it has been indicated that the level of intellectual functioning is a product of the interaction of three variables: native potential, motivation to learn, and environmental stimulus potential. It seems highly probable that a child who is brought up in an environment in which motivation to learn is not instilled, or in which environmental stimulus potential is minimal, or whose parents are of limited innate intelligence, also may well grow up in an environment that predisposes him to a psychiatric condition. In other words, intellectual deprivation is probably frequently associated with emotional deprivation with the same individual subjected to both. It is self-evident that this need not always be the case and that emotional deprivation may occur even under conditions that lead to the greatest development of native intellectual potential. Although it has been assumed that this association occurs as a function of the nature of the environment, the possibility of a hereditary linkage cannot be ruled out. It may be that a genetic factor that limits the development of intelligence frequently occurs in conjunction with a hereditary predisposition to emotional disturbance.

The findings also raise certain questions with regard to intellectual deterioration in the emotionally ill. If those who suffer from functional disorders are on the average of lower intelligence than the population as a whole, and if those who remain ill or have a recurrence are as a group even less intelligent than those who recover relatively quickly, initial intelligence must be controlled in any study aimed at determining whether deterioration actually occurs. Without such control it is impossible to say whether any obtained difference is primarily attributable to a lower level of initial intelligence in the chronic group or basically a reflection of true deterioration. Recent studies (1, 7) have raised serious doubts as to the existence of intellectual deterioration among the older age groups of the population. A re-examination of the deterioration concept as applied to those suffering from chronic emotional disturbance also seems in order.

SUMMARY

AGCT scores at the time of entry into military service were obtained for representative samples of men inducted during late 1942 and prematurely separated for psychosis and psychoneurosis. A comparison of these scores with those for total inductions in the period September 1942 through May 1943 revealed a reliably lower intelligence among both psychiatric groups. However, no reliable difference was found between the neurotics and psychotics. A further comparison between the scores of those prematurely discharged soldiers who were hospitalized for emotional disturbance as civilians and the scores of those not hospitalized indicated that the hospitalized veterans originally had a much lower intelligence level.

REFERENCES

1. BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818.
2. BRILL, N. Q., & BEEBE, G. W. *A follow-up study of war neuroses*. Washington: U. S. Veterans Administration, 1955.
3. DAVENPORT, R. K. Implication of military selection and classification in relation to universal military training. *J. Negro Educ.*, 1946, **15**, 585–594.

4. GINZBERG, E., & BRAY, D. W. *The uneducated.* New York: Columbia Univer. Press, 1953.
5. GINZBERG, E., HERMA, J. L., & GINSBURG, S. W. *Psychiatry and military manpower policy.* New York: Columbia Univer. Press, 1953.
6. MASON, C. F. Pre-illness intelligence of mental hospital patients. *J. consult. Psychol.*, 1956, **20**, 297–300.
7. MINER, J. B. *Intelligence in the United States.* New York: Springer, 1957.
8. ROE, ANNE, & SHAKOW, D. Intelligence in mental disorder. *Ann. N.Y. Acad. Sci.*, 1942, **42**, 361–490.
9. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25**, 555–562.
10. WEISSKOPF, EDITH A. Intellectual malfunctioning and personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 410–423.

Received April 10, 1957.

INFLUENCING ETHNOCENTRISM IN SMALL DISCUSSION GROUPS THROUGH A FILM COMMUNICATION[1]

LEONARD L. MITNICK[2] AND ELLIOTT McGINNIES
University of Maryland

THE experimental literature dealing with persuasive communications has been summarized and given a theoretical orientation by Hovland et al. (6). Many questions, however, remain unanswered, among them problems concerning the relationships between certain aspects of group composition, interaction, and the direction and stability of attitude change under the impact of a persuasive communication. The present paper is concerned with learning and attitude change among the members of small, attitudinally homogeneous groups that experienced varying degrees of participation with respect to discussion of a motion picture film dealing with the problem of group prejudice and conflict.

Motion picture films may exert a significant influence upon the attitudes and opinions of viewers (7, 14, 15, 16). It has also been shown that group participation in decision-making is a more effective technique for generating acceptance than mere exposure to persuasive influences (1, 2, 10). Active participation either imposed by the experimenter or allowed to develop at the initiative of the group members seems to encourage acceptance of the content of a communication (8, 10).

A difficulty in generalizing too confidently from such findings arises when one considers that active rehearsal of this nature is assumed to reinforce the content of the communication and, hence, might be expected to facilitate attitude change only in those situations where the participants concur with the communication content. In cases where the communication tends to counteract existing attitudes in the audience, it is questionable that active participation would provide the kind of conditions that would enhance the effects of the communication. It seems particularly questionable with respect to groups that are unified in their antagonism to the communication content. Prediction of the direction, extent, and stability of attitude change in groups predisposed in different ways toward a controversial film thus poses a problem. What, for example, would be the nature of attitude changes in groups composed homogeneously of individuals who have a positive, negative, or neutral bias toward the content of the communication when opportunity for discussion of the material is extended or withheld? The experimental literature in the field of communication and persuasion provides no basis for making clear-cut predictions under several conditions described.

In one sense, as Hovland, Janis, and Kelly (6) have pointed out, the effectiveness of a persuasive communication is a matter of learning. Any attitude change following exposure to a film, for example, may reflect the amount of new information conveyed by the film to the audience. Such change seems to be most effective with respect to those items that are covered specifically in the film content (7). A relationship between attitude and retention of information has been found by Doob (4), who reports that intensity rather than direction of attitude facilitates retention of the content of a communication. If such is the case, then the familiar principles of learning do not apply without emendation to persuasive communication, since superior retention for pleasant, or acceptable, as opposed to unpleasant, or unacceptable, material has been reported in several studies (5, 17, 18). A further problem intrudes here, namely the effects of initial attitude upon learning among the recipients of a persuasive communication. Does initial attitude influence the amount of learning from a relevant communication, and is attitude change related to such learning?

The problems suggested here become even more complex when one adds the variable of

[1] This research was supported in part by a special grant from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service. The assistance of Dr. Willard Vaughan and Mr. Clagett Smith, our co-researchers, is gratefully acknowledged. Dr. Ray C. Hackman assisted in the design and statistical analysis. This paper is based upon portions of the doctoral dissertation of the first author.

[2] Now with the Human Resources Research Office, Washington, D. C.

active participation to the audience settings, since it seems well-established that participation in a communication situation enhances learning on the part of those who participate (9, 13). The present study was designed to measure the relative influence of the factors of initial attitude and learning upon participation, the retention of information, and the direction, extent, and stability of attitude change.

METHOD

Subjects

Approximately 400 students from two high schools[3] responded to a slightly modified version of the California Ethnocentrism (E) Scale. In each school, 27 students were selected from each extreme and from the middle of the distribution of E scores, which closely approximated a Gaussian curve. The three levels of ethnocentrism represented will be referred to as low E, high E, and middle E, indicating the portion of the score distribution from which the *S*s were chosen. This preliminary testing took place about one month prior to the experimental sessions. Means and variances were determined for the students of the two schools and no significant differences between schools were found for either of these measures.

Procedure

Within each category of ethnocentrism, nine *S*s were assigned at random to each of three experimental conditions, designated as film-discussion, film-alone, and control. This was done within schools, so that a total of 18 groups, or 162 *S*s, representing three degrees of ethnocentrism participated in the experiment. The film-discussion groups first viewed the film "The High Wall," which treats group prejudice as a communicable disease and traces its origins in the family and community. The members of these groups, representing low, high, and intermediate degrees of ethnocentric disposition, were given thirty minutes in which to discuss the film and related issues. The group members were free to participate in the discussion or not, and the discussion leader assumed a permissive, or nondirective, role. The same discussion leader served for all of the groups. The discussions were terminated either at the end of half an hour or when it appeared that the group members had no more comments to offer. At the conclusion of the discussion, the *S*s again answered the E scale.

Members of the film-alone groups viewed the film under comparable conditions but did not engage in discussion of it. They again filled out the E scale upon conclusion of the film showing. The control groups responded to the E scale at the same times as the experimental groups but without any intervening exposure to either the film or a discussion.

In order to determine the amount of learning of film content by the *S*s, a 33-item multiple-choice Information test covering factual material in the film was devised through pretesting and item-analysis. The items on the test covered such points as identification of characters in the film and details of the action in certain scenes. The items did not deal directly with the problem of prejudice. The Information test, therefore, was not a measure of attitude-related material but, rather, was an index of the sheer amount of detail retained from exposure to the film. Pretesting with this form would have been meaningless, since the pattern of responses would necessarily be random for *S*s who had not viewed the film. This test was administered to each of the experimental groups (film-discussion and film-alone) immediately after the second administration of the E scale, following presentation of the film. Both the E scale and the Information test were readministered a final time one month after the film showing in order to determine retention of information and stability of attitude change during this interval. The control groups, of course, did not take the Information test at any time, since it dealt exclusively with the content of the film.

All of the discussions were tape-recorded, and the recording equipment was present whether a discussion was to follow the film or not. None of the apparatus was concealed from the discussion groups, but they were assured that all records would be held in confidence. Two other individuals were present in the experimental situation; they were the "spotters" who, from the rear of the room, recorded participation of an *S* in terms of *S*'s seat by row and seat number. When an individual participated, his location was noted by both spotters in terms of these references. Thus, it was possible to maintain a running account of participation by each *S*, and to identify each comment with its source upon typed, verbatim transcripts of the recordings. This procedure is described in detail elsewhere (11).

RESULTS

Attitude Change

The measure used to assess attitude change was a difference score determined by subtracting the post treatment score from the initial score on the E scale. Since it is possible that a change of position on an attitude scale may be related to one's initial attitude, the correlation coefficient between initial attitudes and difference scores was computed and found to be 0.16, indicating that less than three per cent of the variability in the difference scores could be accounted for by the initial scores.

[3] Miss Rowenetta Allen, Supervisor of Instruction of Prince George's County, Maryland, kindly granted permission for inviting two high schools to participate in this research. Mr. Allan Chotiner and Mr. Fred Wampler of High Point Senior High School, and Mr. Kalman Vozar and Mr. Ralph Angel of Bladensburg Senior High School cooperated generously in all phases of the study.

TABLE 1
ANALYSIS OF VARIANCE OF ATTITUDE CHANGES

Source	*df*	Sums of Squares	Mean Squares	*F*[a]
Predisposition (P)	2	779.4	389.7	3.62
Treatments (T)	2	6506.9	3253.5	30.18**
P × T	4	431.0	107.8	5.36**
Within groups	153	3068.1	20.1	
Total	161	10785.4		

[a] A single asterisk indicates rejection of the null hypothesis at the .05 level of confidence, and the double asterisk indicates rejection at the .01 level.

TABLE 2
MEAN ATTITUDE CHANGE SCORES[a]

	Low E	Middle E	High E
Control	32.5	35.9	37.8
Film-alone	44.9	48.5	54.7
Film-discussion	47.5	49.5	47.6

[a] For ease in computation a constant of +40 was added to all difference scores.

Data from the two high schools were combined for all statistical analyses. After Bartlett's test provided no basis for rejecting the assumption of homogenous variances, an analysis of variance was performed on the attitude change scores. As shown in Table 1, the *F* values for the treatment × predisposition interaction and treatment as a main effect were significant at the .01 level of confidence. Table 2 shows the mean values for each treatment and predisposition. Among the high E individuals differences in attitude change between the film-alone and film-discussion conditions were significant at the .05 level of confidence. The differences between the film-alone and film-discussions treatment for the middle E and low E groups were not significant. For the film-alone condition the high E groups showed significantly more change than the other two groups, but for the film-discussion treatment there were no significant differences among predisposition groups. It is evident, then, that giving the high E groups the opportunity to discuss the film counteracted some of the effects of the film, while for the low E groups a reversal of this trend was noted. The significance of treatment as a main effect was due to differences between control groups and both experimental groups; there was no significant difference between the combined film-discussion groups and the film-alone groups.

Fig. 1 presents the mean values for the initial, posttreatment and delayed administration of the E scale for all treatments. Since predisposition was not associated with attitude change, the various predisposition groups are combined according to treatment. The attitude change scores (initial-minus-post) for the film-discussion groups and for the film-alone groups both differed significantly from the control groups at the .01 level of confidence, but the experimental groups did not differ with respect to attitude change. In short, significant reductions in personal ethnocentrism were brought about by both the film-discussion and film-alone treatments, while the control groups showed no change in ethnocentric attitudes.

*S*s were retested with the E scale one month after treatment, and an analysis of variance was performed on the differences between the initial scores and the delayed scores on the E scale. Inasmuch as some of the *S*s were absent at the time of the delayed E scale, their scores were estimated by a procedure suggested by Cochran and Cox (3). Before performing this analysis, the correlation coefficient between this difference score and the initial score was determined and found to be 0.04, indicating that only 0.16% of the variability in the difference scores could be attributed to the initial

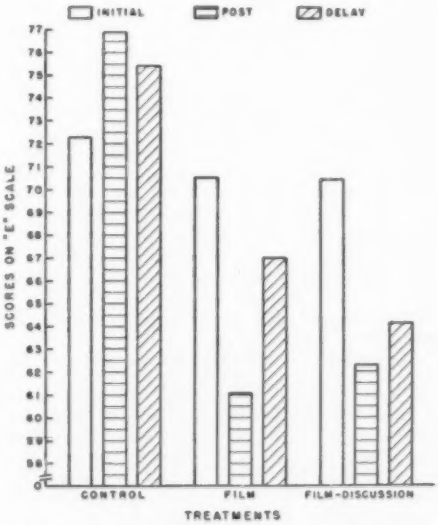

FIG. 1. INITIAL, POSTTREATMENT, AND DELAY SCORES ON THE "E" SCALE FOR EACH TREATMENT

TABLE 3
ANALYSIS OF VARIANCE OF STABILITY OF ATTITUDE CHANGE

Source	*df*	Sums of Squares	Mean Squares	*F*
Predisposition (P)	2	40.1	20.0	—
Treatments (T)	2	2563.3	1281.6	9.52**
P × T	4	625.3	156.3	1.16
Within groups	139	18709.7	134.6	
Total	147	21938.4		

scores. An examination of this analysis, shown in Table 3, shows that the obtained *F* value for treatment was significant at the .01 level of confidence. Fig. 1 presents the delay scores for each treatment. Differences between the retention of attitude change for the film-alone and film-discussion groups differed significantly from the controls at the .01 level of confidence, but these two groups did not differ from each other when compared with respect to initial score minus delayed score.

A further comparison of the scores represented in Fig. 1 was made. Within each experimental condition, the scores on the posttreatment and delayed administration of the E scale were compared. Neither the control groups nor the film-discussion groups differed significantly on these two measures, indicating that for these two conditions there was no significant change one month after treatment from posttreatment scores. However, the film-alone groups showed a significant change on these two measures. At the time of the delayed measurement, the film-alone groups had moved significantly from their posttreatment score position in the direction of their pre-experimental position.

It was previously indicated that on the posttreatment measure the film-alone groups and the film-discussion groups did not differ significantly. A breakdown of the film-discussion groups was made into active and passive participants, with the active participants defined as those individuals who made one or more comments in the discussion regardless of length, and the passive participants defined as those individuals who made no contribution to the discussion. Of the 54 *S*s who comprised the film-discussion groups, 37 were active participants and 17 were passive participants. Fig. 2 shows the mean values for three degrees of participation on each administration of the E scale. None of the changes from initial to posttreatment scores differed significantly from one another.

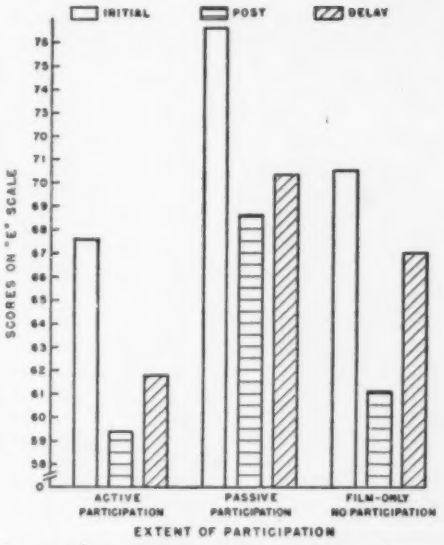

FIG. 2. INITIAL, POSTTREATMENT, AND DELAY SCORES ON THE "E" SCALE FOR EACH LEVEL OF PARTICIPATION

A comparison of the posttreatment and delayed E-scale positions had shown that the film-alone groups moved significantly from their posttreatment scores in the direction of their original position, while the control and film-discussion groups showed no such change. Is this difference attributable to greater attitude change among the active participants in the film-discussion groups, or does it result from the process of discussion influencing all members present, whether active or passive? In order to provide an answer to this question, the active and passive participants were compared on their posttreatment and delayed scores. Fig. 2 shows these differences graphically. The differences between the posttreatment and delayed E scores were not significant for either active or passive participants. That is, neither the active nor the passive discussion participants changed significantly from their posttreatment measures after an interval of one month. The film-alone groups on the other hand, regressed toward their earlier position. The stability of attitude change characterizing both the active and passive participants suggests that presence during a discussion is as

effective as active participation in generating greater permanency of attitude change.

Learning

The relationship of amount of information learned from the film to the major independent variables of the experiment was also evaluated by analysis of variance. Since the relationship between learning and participation, however, was ascertainable only in those groups that held discussions, that analysis is presented separately.

Examination of Table 4 shows that the *F* value for predisposition was significant at the .01 level of confidence. The mean scores for the three predisposition groups were: low E, 24.8; middle E, 22.3; high E, 19.8. All differences between these means were significant at the .01 level of confidence, indicating an inverse relationship between initial degree of ethnocentrism and learning of information specific to the film. Since treatments were not a significant source of variability, it is apparent that discussion following the film communication did not facilitate recall of details from the film.

TABLE 4

ANALYSIS OF VARIANCE OF SCORES ON THE INFORMATION TEST

Source	*df*	Sums of Squares	Mean Squares	*F*
Predisposition (P)	2	450.7	225.4	15.65**
Treatment (T)	1	16.3	16.3	1.13
P × T	2	18.7	9.4	
Within groups	102	1468.0	14.4	
Total	107	1953.7		

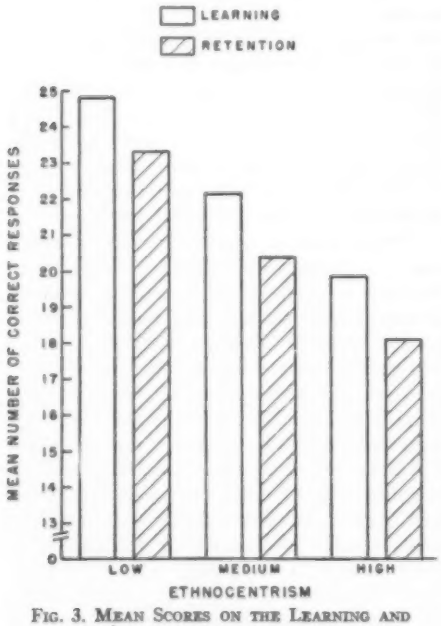

FIG. 3. MEAN SCORES ON THE LEARNING AND RETENTION TEST FOR EACH PREDISPOSITION GROUP

Fig. 3 presents the mean learning and retention scores of each predisposition group. It is evident that individuals who scored high on the E scale learned the least from the film, while those who were low on the E scale learned the most from the film. The middle E groups fell in between these extremes. Those individuals, in short, who were assumed on the basis of the E scores to be favorably disposed toward the content of the communication learned significantly more factual material from it than did those *S*s who were antagonistic to the film's theme, even though the information learned was not directly related to the topic of prejudice. That this relationship was not confounded with differences in intelligence among the *S*s was attested by the fact that the correlation between E scores and scores on the California Mental Maturity Test was not significant.

An analysis of variance was also performed on the retention of information scores to determine whether the experimental variables were reflected in differential retention of film content among the *S*s one month after the first testing. The scores of the absentees were again estimated by the procedure suggested by Cochran and Cox (3). Results of the analysis of variance which are given in Table 5 show that predisposition is significant at the .01 level. The differences in retention scores among the three predisposition conditions are presented graphically in Fig. 3. An increased degree of ethnocentrism, in general, is associated with a greater decrement in retention of information. It should be noted, however, that a correlation of .73 ($P < .01$) was found between learning and retention, so that the differences in retention of information may be ascribed to amount of material initially learned rather than to predisposition as such.

Finally, an effort was made to determine

TABLE 5
ANALYSIS OF VARIANCE OF INFORMATION RETENTION SCORE

Source	*df*	Sum of Squares	Mean Squares	*F*
Predisposition (P)	2	483.6	241.8	15.80**
Treatment (T)	1	35.6	35.6	—
P × T	2	90.7	45.4	2.97
Within groups	94	1436.5	15.3	
Total	99	2046.4		

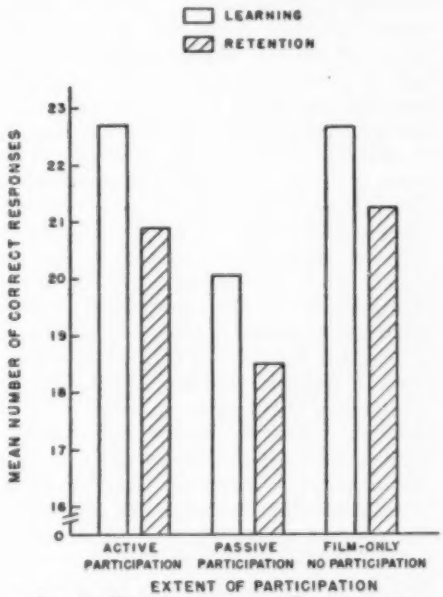

FIG. 4. MEAN LEARNING AND RETENTION OF INFORMATION SCORES FOR EACH LEVEL OF PARTICIPATION

whether those individuals who participated voluntarily in the discussions learned more from the film than those who did not participate. Fig. 4 shows the mean learning scores made by the active participants, passive participants, and film-alone *S*s. Evaluation by means of *t* tests revealed no significant differences between the active participants and members of the film-alone groups. Both the active participants and the film-alone *S*s, however, learned significantly more from the film than the passive discussion participants. This advantage persisted on the retention measure one month later.

In an attempt to determine whether attitude change was related to the amount of information learned, a correlation was computed between these two variables. The obtained value of .03 indicated that there was no significant relationship between learning of film details and attitude change.

DISCUSSION

Attitude change in response to a persuasive sound film was a joint function of opportunity for discussion of the film and initial attitude. Predisposition alone was not related to attitude change. It had been expected that in the low E groups a greater effect might be obtained with those members who held discussions, since these individuals had views that were essentially in accord with those contained in the film. (Even the low Es could show attitude change since the mean score of these *S*s was 26.7 points below the ceiling of the E scale.) The discussion might reasonably have been expected to reinforce the initial attitudes of these *S*s and thus facilitate a greater change in attitude than that produced among the low Es who did not hold discussions. In the high E groups, on the other hand, a possibility existed that expression of opinions at variance with those presented in the film might mitigate the film's effect, and that the nondiscussion groups might therefore show greater attitude change. Among the neutral or middle E *S*s, no difference was expected between the film and film discussion conditions.

These hypotheses were generally supported by the data. Although the difference in attitude change in the low E groups under film-alone and film-discussion conditions was not significant, the trend was in the expected direction. The high Es showed significantly less change under discussion conditions. In general, it may be suggested that discussion of a persuasive communication does facilitate attitude change among those initially favorable to the communication content. On the other hand, discussion of a communication by homogeneous groups that are antagonistic to the content detracts from the effectiveness of the communication. This result may be attributed to the fact that discussion provides an opportunity for such individuals to express their initial views and thus partially to negate the persuasive effects that have been generated. Examination of the discussion transcripts supports this interpretation. The high Es devoted

most of the discussion periods to expression of their antipathies toward Negroes, while the low E and middle E groups tended to examine the general problems of group prejudice raised by the film. The middle *S*s having no clear-cut opinion either pro or con held very apathetic discussions and, therefore, neither reinforced nor mitigated the effects of the film.

While there were no differences in attitude change between the film-alone and film-discussion groups immediately following the communication, the film-alone groups regressed significantly from their posttreatment positions when measured one month later; the film-discussion groups and the controls showed no such change. The discussions probably gave those *S*s present an opportunity to learn how other members of their group perceived the film and permitted the convergence of views toward a common norm. Apparently influenced by such a consensus effect, the film-discussion groups maintained their posttreatment attitude change. The film-alone groups, not having experienced this interaction, regressed towards their initial scores on the delayed measurement. Thus, participation in the form of group discussion following a communication may have some value in maintaining attitude changes produced by the communication, even when the initial gains are not measurably affected. Our results indicated that extent of participation was not a factor in determining initial attitude change. Participation, however, did influence the permanence of posttreatment attitude changes. On the delayed measurement neither the active nor the passive participants regressed significantly from their posttreatment scores, while the film-alone groups did show such a regression. Stability of the attitude change, therefore, cannot be attributable to active participation alone, but must be due in part to mere presence in a group discussion. Although active participation has been defined as verbal commitment, the possibility of nonverbal communication during the discussion must be considered. Reinforcement of opinion change may be the result of a number of aspects of the group discussion situation not limited to verbal interchange.

A decreasing amount of factual material was learned from the film in moving from the groups low in ethnocentrism to those high in ethnocentrism. This finding implies that those *S*s whose attitudes were assumed to be congruent with the content of the film learned the most from it, while the *S*s whose attitudes were in conflict with the film content learned the least amount of factual material. Our findings thus substantiate the hypothesis that learning from a persuasive communication is in part a function of initial attitude toward the communication content.

A content analysis of the discussion materials indicated that discussion in the low E groups tended to center about the arguments presented by the film, whereas discussion in the high E groups departed almost completely from the film. While it was not determined how much information in the film was reviewed in each discussion, the fact that more comments in the low E groups were related to the film increased the probability that specific informational items would be mentioned. Actually, the effects of discussion content and initial attitude are confounded in the discussion groups. In the film-alone groups, however, no such confounding of these two variables is present, and here the low Es also learned significantly more from the film than the high Es. In addition, the analysis of variance of Information test scores showed that treatment (discussion vs. no discussion) was not a significant source of variability. These considerations indicate that discussion of a communication does not necessarily alter the amount of information acquired from the communication. In the present instance, the discussion comments were not such as to involve rehearsal of factual content of the film. The probability that the amount of practice furnished by the discussion was slight would account for the lack of differences in learning between the film-discussion and film-alone groups.

Although there were no significant differences on the initial administration of the E scale between the active and passive participants, a significant difference was observed between these groups in amount of information learned. The fact that the active participants learned more than the passive participants agrees with the findings of Kurtz and Hovland (9). In light of this, one might predict that the film-alone *S*s, having no opportunity for participation, would learn less from the film than either the active or passive members of the discussion groups. However, the results

showed that the film-alone *S*s learned significantly more than the passive participants, and, in fact, did not differ significantly from the active participants. These findings seem to indicate that participation depended on amount of material learned, since the film-alone groups undoubtedly contained some individuals who would have participated actively in discussion given the opportunity. The presence of these potential discussants would tend to raise the average learning score in the film-alone groups above the mean of the passive members of the discussion groups. In other words, the mean learning score for the film-alone groups represents the joint contribution of potentially active and passive discussion participants.

Data reported by McGinnies and Vaughan (12) support the contention that participation in a free discussion depends upon the amount of relevant information possessed by *S*s. Their results indicated that the *S*s who participated in discussions following mental health films rated themselves higher on a self-rating scale of mental health information than those who did not participate in the discussion. They also reported that the discussants had seen significantly more mental health films than the nondiscussants. Participation in a discussion appears to depend in part upon knowledge of the problem.

At the time of the delayed administration of the Information test, the low E groups obtained the highest scores, and the high E groups achieved the poorest retention scores. Although predisposition, or initial attitude, was not a significant source of variance when the retention scores were adjusted for initial learning, the predisposition × treatment interaction was nearly significant. In other words, the observed differences in retention of information among the several predisposition groups probably depended upon the occurrence or nonoccurrence of a discussion following the film. For the low E and high E groups, the more effective treatment was the film-discussion situation.

Retention of information was obviously unrelated to extent of participation in discussion, since the film-alone groups achieved retention scores equal to those of the active discussion participants and higher than those of the passive participants. Furthermore, no significant differences existed between the active and passive participants in this respect.

SUMMARY AND CONCLUSIONS

This study investigated the effects of ethnocentric attitudes and discussion participation upon the degree and stability of learning and attitude change in response to a persuasive communication. Twelve experimental groups, each composed of nine high school students, viewed a sound film "The High Wall." Six of the groups held discussions of the film, while the remaining six did not. The groups were formed on the basis of scores on the California Ethnocentrism Scale, so that two groups under each of the experimental conditions represented low, middle, and high scores on the E scale. Following experimental treatment, film-discussion or film-alone, all of the group members were tested for attitude change and amount of factual information learned from the film. Six control groups were given pre- and posttests on the questionnaires without intervening exposure to either the film or discussion. One month following the experimental treatments, all groups were retested on the E scale, and the experimental groups were also readministered the Information test.

The major findings of this experiment may be summarized as follows:

1. Significant reductions in ethnocentrism, as measured by the E scale, were produced in both the film-discussion and film-alone groups. For prejudiced *S*s, the effect was significantly less in the discussion groups than in the nondiscussion groups. Ethnocentric predisposition alone, however, did not affect the extent of attitude change induced by the film.

2. Stability of attitude change was a function of experimental treatment. The active and passive members of discussion groups, when tested one month later, had largely retained their posttreatment attitude gains, while the film-alone groups had regressed significantly toward their original attitudes.

3. Amount of information learned from the film was related to initial attitude. Those individuals who were low in ethnocentrism learned more from the film than those high in ethnocentrism. This trend persisted in the retention measures taken a month later.

4. Active discussion participants learned more from the film than passive participants.

Analysis of this finding led to the conclusion that extent of participation in discussion depends upon the amount of learning that has taken place.

REFERENCES

1. BAVELAS, A. Role playing and management training. *Sociatry*, 1947, **1**, 183–191.
2. COCH, L., & FRENCH, R. P., JR. Overcoming resistances to change. *Hum. Relat.*, 1948, **1**, 512–532.
3. COCHRAN, W. G., & COX, G. M. *Experimental design* New York: Wiley, 1950.
4. DOOB, L. W. Effects of initial serial positions and attitude upon recall conditions of low motivation. *J. abnorm. soc. Psychol.*, 1953, **48**, 199–205.
5. EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–50.
6. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
7. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.
8. JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.
9. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.
10. LEWIN, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1947. Pp. 330–344.
11. McGINNIES, E. A method for matching anonymous questionnaire data with group discussion material. *J. abnorm. soc. Psychol.*, 1956, **52**, 139–140.
12. McGINNIES, E., & VAUGHAN, W. Some biographical determiners of participation in group discussion. *J. appl. Psychol.*, 1957, **41**, 179–185.
13. MICHAEL, D. M., & MACCOBY, N. Factors influencing verbal learning from films under varying conditions of audience participation. *J. exp. Psychol.*, 1953, **46**, 411–418.
14. PETERSON, R. C., & THURSTONE, L. L. *Motion pictures and the social attitudes of children*. New York: Macmillan, 1933.
15. RAMSEYER, L. L. A study of the influence of documentary films on social attitudes. Unpublished doctoral dissertation, Ohio State Univer., 1938.
16. RATHS, L. E., & TRAGER, F. N. Public opinion and crossfire. *J. educ. Soc.*, 1948, **21**, 345–368.
17. SHARP, A. A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, **22**, 395–418.
18. WATSON, W. S., & HARTMAN, G. W. The rigidity of a basic attitudinal frame. *J. abnorm. soc. Psychol.*, 1939, **34**, 314–335.

Received April 15, 1957.

SOME FACTORS IN SELF-DISCLOSURE[1]

SIDNEY M. JOURARD[2] AND PAUL LASAKOW

University of Alabama Medical Center and University of Alabama Birmingham Center

THE present paper describes a questionnaire method for measuring the amount and content of self-disclosure to selected "target-persons," and reports the results of three exploratory studies. Self-disclosure refers to the process of making the self known to other persons; "target-persons" are persons to whom information about the self is communicated.

The process of self-disclosure has been studied by others from various points of view. Block (1) and Block and Bennett (2) have demonstrated that the content of communication about the self is a function of variations in "own-role." Lewin (7) noted differences between typical Germans and Americans regarding their readiness to confide personal information to others. Jourard (6) has suggested that accurate portrayal of the self to others is an identifying criterion of healthy personality, while neurosis is related to inability to know one's "real self" and to make it known to others. Characterological studies of Fromm (3), Riesman (9), and Horney (4) have called attention to a tendency common among persons in our society, to misrepresent the self to others. This tendency is central to the "marketing personality," the "other-directed character," and the "self-alienated" individual, as these have been described by their respective authors. Since much of social science is founded upon the self-disclosures of respondents, the conditions and dimensions of self-disclosure bear directly upon the validity of many purported facts in the social sciences.

From the foregoing, it may be concluded that systematic analysis of self-disclosure holds promise of yielding information that is relevant to diverse areas of theory and method.

The following questions were proposed for investigation:

1. Do subjects (*Ss*) vary in the extent to which they disclose themselves to different target-persons, for example, mother, father, male friend, and female friend? What is the effect of the *Ss*' marital status on self-disclosure to parents and friends? What is the effect of the *Ss*' feelings and attitudes toward particular target-persons upon self-disclosure to them? The last question was investigated only with respect to the relationship between *Ss*' disclosure of self to parents, and their feelings and attitudes toward their parents.

2. Are there differences between categories of information about the self (*aspects* of self) with respect to self-disclosure? Do *Ss* tend to disclose some aspects of self more fully than others?

3. Are there differences ascertainable between Negro and white *Ss* with respect to self-disclosure?

4. Are there sex differences regarding self-disclosure?

METHOD

Instruments

The Self-disclosure Questionnaire. A sixty-item questionnaire was devised. As can be seen in Table 1, the items are classified in groups of ten within each of six more general categories of information about the self (*aspects*). *Ss* were given the following instructions for completing the questionnaire:

The answer-sheet which you have been given has columns with the headings "Mother," "Father," "Male Friend," "Female Friend," and "Spouse." You are to read each item on the questionnaire, and then indicate on the answer-sheet the extent that you have talked about that item to each person; that is, the extent to which you have made yourself known to that person. Use the rating-scale that you see on the answers sheet to describe the extent that you have talked about each item.

The self-disclosure rating-scale was as follows:

0: Have told the other person nothing about this aspect of me.

1: Have talked in general terms about this item. The other person has only a general idea about this aspect of me.

2: Have talked in full and complete detail about this item to the other person. He knows me fully in this respect, and could describe me accurately.

X: Have lied or misrepresented myself to the other person so that he has a false picture of me.

The numerical entries were summed (*X*'s were counted

[1] The authors are indebted to Drs. A. J. Riopelle, Emory University, and O. Lacey, University of Alabama, for statistical advice.

[2] Now at the University of Florida, Gainesville.

TABLE 1

THE SELF-DISCLOSURE QUESTIONNAIRE

Attitudes and opinions

1. What I think and feel about religion; my personal religious views.
2. My personal opinions and feelings about other religious groups than my own, e.g., Protestants, Catholics, Jews, atheists.
3. My views on communism.
4. My views on the present government—the president, government policies, etc.
5. My views on the question of racial integration in schools, transportation, etc.
6. My personal views on drinking.
7. My personal views on sexual morality—how I feel that I and others ought to behave in sexual matters.
8. My personal standards of beauty and attractiveness in women—what I consider to be attractive in a woman.
9. The things that I regard as desirable for a man to be—what I look for in a man.
10. My feelings about how parents ought to deal with children.

Tastes and interests

1. My favorite foods, the ways I like food prepared, and my food dislikes.
2. My favorite beverages, and the ones I don't like.
3. My likes and dislikes in music.
4. My favorite reading matter.
5. The kinds of movies that I like to see best; the TV shows that are my favorites.
6. My tastes in clothing.
7. The style of house, and the kinds of furnishings that I like best.
8. The kind of party, or social gathering that I like best, and the kind that would bore me, or that I wouldn't enjoy.
9. My favorite ways of spending spare time, e.g., hunting, reading, cards, sports events, parties, dancing, etc.
10. What I would appreciate most for a present.

Work (or studies)

1. What I find to be the worst pressures and strains in my work.
2. What I find to be the most boring and unenjoyable aspects of my work.
3. What I enjoy most, and get the most satisfaction from in my present work.
4. What I feel are *my* shortcomings and handicaps that prevent me from working as I'd like to, or that prevent me from getting further ahead in my work.
5. What I feel are my special strong points and qualifications for my work.
6. How I feel that my work is appreciated by others (e.g., boss, fellow-workers, teacher, husband, etc.)
7. My ambitions and goals in my work.
8. My feelings about the salary or rewards that I get for my work.
9. How I feel about the choice of career that I have made—whether or not I'm satisfied with it.
10. How I really feel about the people that I work for, or work with.

Money

1. How much money I make at my work, or get as an allowance.
2. Whether or not I owe money; if so, *how much.*

TABLE 1—*Continued*

3. Whom I owe money to at present; or whom I have borrowed from in the past.
4. Whether or not I have savings, and the amount.
5. Whether or not others owe me money; the amount, and who owes it to me.
6. Whether or not I gamble; if so, the way I gamble, and the extent of it.
7. All of my present sources of income—wages, fees, allowance, dividends, etc.
8. My total financial worth, including property, savings, bonds, insurance, etc.
9. My most pressing need for money right now, e.g., outstanding bills, some major purchase that is desired or needed.
10. How I budget my money—the proportion that goes to necessities, luxuries, etc.

Personality

1. The aspects of my personality that I dislike, worry about, that I regard as a handicap to me.
2. What feelings, if any, that I have trouble expressing or controlling.
3. The facts of my present sex life—including knowledge of how I get sexual gratification; any problems that I might have; with whom I have relations, if anybody.
4. Whether or not I feel that I am attractive to the opposite sex; my problems, if any, about getting favorable attention from the opposite sex.
5. Things in the past or present that I feel ashamed and guilty about.
6. The kinds of things that just make me furious.
7. What it takes to get me feeling real depressed and blue.
8. What it takes to get me real worried, anxious, and afraid.
9. What it takes to hurt my feelings deeply.
10. The kinds of things that make me especially proud of myself, elated, full of self-esteem or self-respect.

Body

1. My feelings about the appearance of my face—things I don't like, and things that I might like about my face and head—nose, eyes, hair, teeth, etc.
2. How I wish I looked: my ideals for overall appearance.
3. My feelings about different parts of my body—legs, hips, waist, weight, chest, or bust, etc.
4. Any problems and worries that I had with my appearance in the past.
5. Whether or not I now have any health problems—e.g., trouble with sleep, digestion, female complaints, heart condition, allergies, headaches, piles, etc.
6. Whether or not I have any long-range worries or concerns about my health, e.g., cancer, ulcers, heart trouble.
7. My past record of illness and treatment.
8. Whether or not I now make special efforts to keep fit, healthy, and attractive, e.g., calisthenics, diet.
9. My present physical measurements, e.g., height, weight, waist, etc.
10. My feelings about my adequacy in sexual behavior—whether or not I feel able to perform adequately in sex-relationships.

as zeros), yielding totals which constituted the self-disclosure scores.

Seventy white unmarried college students of both sexes were tested for self-disclosure to Mother, Father, Male Friend, and Female Friend, in a study of reliability. Since the questionnaire included 60 items, and there were four target-persons, a total of 240 entries were made by each *S*. These 240 entries were divided into halves by the odd-even method, and the subtotal sums were correlated with each other. The resultant *r*, corrected, was .94, indicating that the *S*s were responding consistently to the questionnaire over all target persons, and all aspects of self.

Parent-Cathexis Questionnaires. Mother-cathexis and Father-cathexis questionnaires, fully described elsewhere (5), were employed to test *S*s' feelings toward their parents. The *S*s rated their feelings about 40 parental traits, e.g., *sense of humor*, *temper*, *ability to make decisions*, in accordance with the following scale:

1: Have strong positive feelings; like very much.
2. Have moderate positive feelings.
3. Have no feelings one way or the other.
4. Have moderate negative feelings.
5. Have strong negative feelings; dislike very much.

High scores indicated negative feelings toward the parents, while low scores signified positive feelings.

Subjects

The *S*s included in the studies to be reported were taken from larger samples drawn from three Alabama college populations: two white liberal arts colleges, a Negro liberal arts college, and a school of nursing located at a medical school. For the combined sample 300 white and Negro liberal arts college sophomores and juniors were obtained, and 55 white nursing students. All *S*s were tested in groups by an examiner of the same race.

For the purpose of analysis, the following randomly selected subsamples were drawn from the combined samples:

1. From the 300 liberal arts students, a subsample of 10 white males, 10 white females, 10 Negro males, and 10 Negro females was drawn for the study of differences in self-disclosure associated with race, sex, targets, and aspects of self. All *S*s were unmarried, and in all cases, the parents were living. Mean ages were: white males, 21.70, *SD* 2.00; white females, 20.30, *SD* .90; Negro males, 22.10, *SD* 2.02; and Negro females, 20.40, *SD* .45. All *S*s had been tested for self-disclosure to Mother, Father, Male Friend, and Female Friend.

2. From all the white respondents in the combined sample, a subsample of 10 married male and 10 married female *S*s was drawn for comparison with the first subsample of 10 unmarried males and 10 unmarried females to test the effects of marriage on self-disclosure patterns. These *S*s had indicated self-disclosure to Mother, Father, Same-sex Friend, and Spouse. Mean age for the married males was 23.40, *SD* 1.43, and for the married females, 20.60, *SD* 2.42.

3. Thirty-one unmarried nursing students comprised the third sample, used to examine the relationship between parent-cathexis and self-disclosure to the parents. Mean age for this group was 18.59, *SD* 3.53.

The data were analyzed according to Lindquist's

TABLE 2

ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF WHITE AND NEGRO, MALE AND FEMALE SUBJECTS

Source	Mean Square	*df*	Error Term	*F*
Between groups	892.33	3	Error (b)	6.78***
Males vs. Females	573.50	1		4.36*
Whites vs. Negroes	2076.82	1		15.78***
Sex × Race	126.67	1		
Error (b)	131.61	36		
Total *df* Between		39		
(Within groups comparisons)				
Between Target-Persons	557.06	3	$Error_1$ (w)	10.89***
Between Aspects of Self	952.01	5	$Error_2$ (w)	40.39***
Target × Aspect	87.49	15	$Error_3$ (w)	15.38***
Group × Target	128.11	9	$Error_1$ (w)	2.50*
Group × Aspect	32.16	15	$Error_2$ (w)	1.36
Group × Target × Aspect	9.61	45	$Error_3$ (w)	1.69**
Total *df* Within		92		
$Error_1$ (w)	51.17	108		
$Error_2$ (w)	23.57	180		
$Error_3$ (w)	5.69	540		
Total Error *df*		828		
Total *df*		959		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

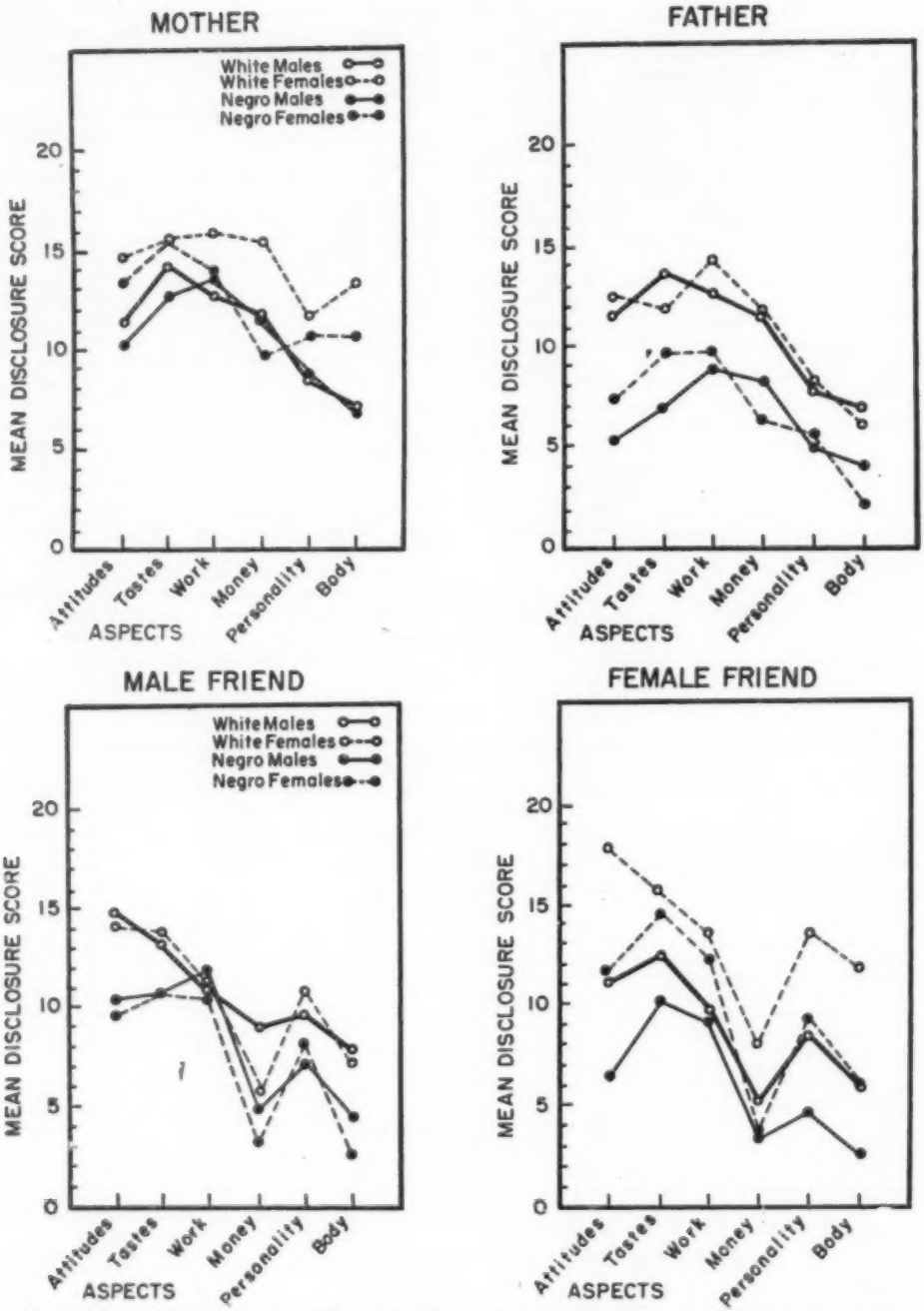

FIG. 1. MEAN SELF-DISCLOSURE OF WHITE AND NEGRO, MALE AND FEMALE *Ss* TO MOTHER, FATHER, MALE FRIEND, AND FEMALE FRIEND

(8) Type VI model for analysis of variance with mixed "between-within" effects. Critical differences for t ratios at the .01 level were computed between all groups, targets, and aspects of self when F ratios proved significant. Pearsonian r's were computed between Mother-cathexis scores and scores for disclosure to the mother, and between Father-cathexis scores and the corresponding disclosure scores, within the group of 31 nursing students to whom these two instruments had been administered.

TABLE 3

DIFFERENCES BETWEEN GROUPS IN MEAN DISCLOSURE SCORES

Group	Mean*	*SD*	Differences**		
			Negro Males	White Females	Negro Females
White males	248.50	68.60	+62.60	−45.10	+33.50
Negro males	185.90	36.10		−107.70	−19.10
White females	293.60	45.20			+78.60
Negro females	215.00	57.60			

* Highest possible score, 480.

** Critical difference for $t = 2.72$, $p < .01$, $df = 36$ is 13.95 (cf. 8, p. 93.). All differences are thus significant at $p < .01$.

TABLE 4

DIFFERENCES BETWEEN TARGETS IN MEAN DISCLOSURE SCORES

Target	Mean*	*SD*	Differences**		
			Father	Male Friend	Female Friend
Mother	72.30	19.50	+20.60***	+17.12***	+16.72***
Father	51.70	24.13		−3.48	−4.88***
Male friend	55.18	22.43			1.40
Female friend	56.58	27.70			

* Highest possible score, 120.

** Critical difference for $t = 2.63$, $p < .01$, $df = 108$ is 4.21 (cf. 8, p. 93.)

*** $p < .01$.

RESULTS

Influence of Race, Sex, Target-Differences, and Aspects of Self

Table 2 shows the results of analysis of variance of the self-disclosure scores of the 40 white and Negro *S*s, while Fig. 1 portrays mean self-disclosure scores classified by targets, groups, and aspects of self. The findings may be summarized as follows:

1. The four groups differed in total self-disclosure. Table 3 shows that the white *S*s disclosed more than the Negroes, and the females more than the males.

2. The combined group of 40 *S*s varied in amount of self-disclosure to different target-persons. They disclosed the most to Mother, and in lesser amount to Father, Male Friend, and Female Friend, as shown in Table 4. Sex- and race-differences in disclosure to the four target-persons is shown clearly in Fig. 1. Noteworthy is the consistently lower amounts of self-disclosure to Father on the part of the Negro *S*s.

3. The combined group of 40 *S*s varied in self-disclosure according to aspects of self. Table 5 shows that two clusters of aspects appeared—a "high disclosure" cluster comprised of Tastes and Interests, Attitudes and Opinions, and Work, and a "low disclosure" cluster that included Money, Personality, and Body.

4. There was significant interaction between targets and aspects, groups and targets, and groups, targets, and aspects. The group-by-aspect interaction was not significant.

TABLE 5

DIFFERENCES BETWEEN ASPECTS OF SELF IN MEAN DISCLOSURE SCORES

Aspect	Mean*	*SD*	Differences**				
			Tastes	Work	Money	Pers.	Body
Att.	45.35	14.20	−4.93***	−2.30	+13.12***	+11.30***	+19.15***
Tastes	50.28	13.98		+2.63	+18.05***	+16.23***	+24.08***
Work	47.65	13.65			+15.42***	+13.60***	+21.45***
Money	32.23	15.65				− 1.82	+ 6.03***
Pers.	34.05	13.45					+ 7.85***
Body	26.20	14.28					

* Highest possible score, 80.

** Critical Difference for $t = 2.63$, $p < .01$, $df = 180$, is 2.87 (cf. 8, p. 93).

*** $p < .01$.

Influence of Marriage

Separate analyses of variance compared married with unmarried males, and married with unmarried females. The results are shown in Table 6, for males, and Table 7, for females. In the analyses, Opposite-sex Friend and Spouse were treated as equivalent target-persons.

No differences were found between married and unmarried *Ss* in total amount of self-disclosure. Figs. 2 and 3 show, however, that: (*a*) married *Ss* disclosed less to the parents and the same-sex friend than unmarried *Ss*, and (*b*) there was more disclosure to the Spouse than to any other target-person by married or unmarried *Ss*.

Marriage thus appears to have the effect, not of increasing or decreasing the total extent to which *Ss* disclose themselves, but of producing a redistribution of self-disclosure. The married *Ss* "concentrated" self-disclosure

TABLE 6

ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF MARRIED AND UNMARRIED MALES

Source	Mean Square	*df*	Error Term	*F*
Between groups	37.95	1	Error (b)	.17
Error (b)	228.81	18		
Total *df* Between		19		
(Within groups comparisons)				
Between Target-Persons	232.99	3	$Error_1$ (w)	6.10**
Between Aspects of Self	477.49	5	$Error_2$ (w)	24.80***
Target × Aspect	21.72	15	$Error_3$ (w)	4.19***
Group × Target	676.95	3	$Error_1$ (w)	17.72***
Group × Aspect	1.00	5	$Error_2$ (w)	.00
Group × Target × Aspect	22.24	15	$Error_3$ (w)	4.49***
Total *df* Within		46		
$Error_1$ (w)	38.21	54		
$Error_2$ (w)	19.25	90		
$Error_3$ (w)	4.95	270		
Total Error *df*		414		
Total *df*		479		

** $p < .01$.
*** $p < .001$.

TABLE 7

ANALYSIS OF VARIANCE OF SELF-DISCLOSURE SCORES OF MARRIED AND UNMARRIED FEMALES

Source	Mean Square	*df*	Error Term	*F*
Between Groups	186.25	1	Error (b)	1.38
Error (b)	135.27	18		
Total *df* between		19		
(Within groups comparisons)				
Between Target-Persons	349.90	3	$Error_1$ (w)	6.08**
Between Aspects of Self	345.54	5	$Error_2$ (w)	18.91***
Target × Aspect	46.64	15	$Error_3$ (w)	8.00***
Group × Target	531.48	3	$Error_1$ (w)	9.23***
Group × Aspect	15.63	5	$Error_2$ (w)	.86
Group × Target × Aspect	26.01	15	$Error_3$ (w)	4.46***
Total *df* Within		46		
$Error_1$ (w)	57.56	54		
$Error_2$ (w)	18.27	90		
$Error_3$ (w)	5.83	270		
Total Error *df*		414		
Total *df*		479		

** $p < .01$.
*** $p < .001$.

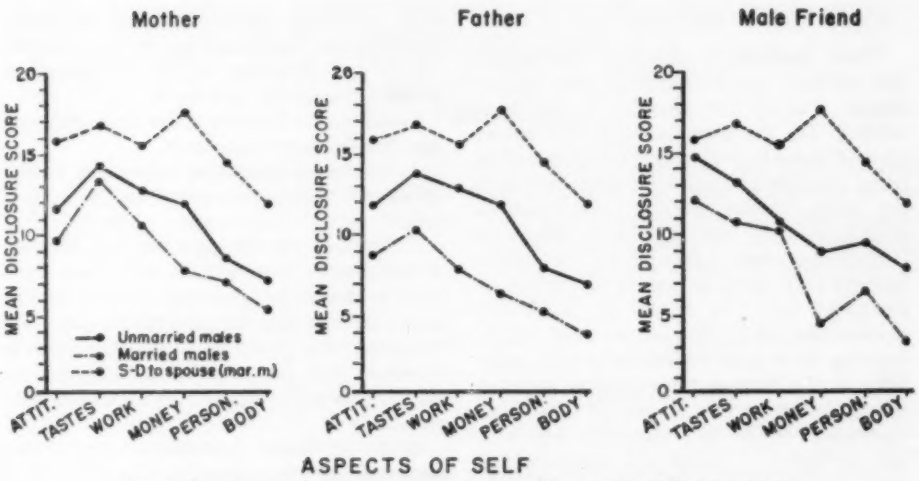

FIG. 2. COMPARISON TO MARRIED AND UNMARRIED MALES ON SELF-DISCLOSURE TO MOTHER, FATHER, AND MALE FRIEND
(Curve for disclosure to Spouse is shown in each panel)

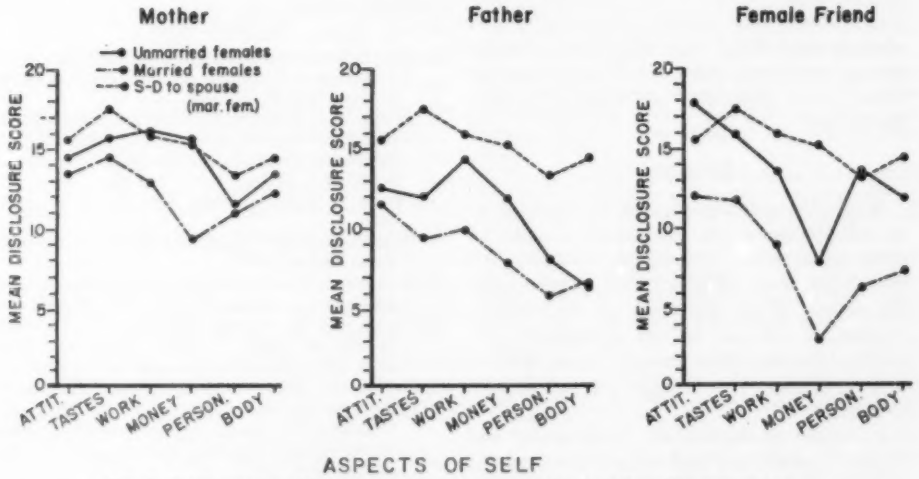

FIG. 3. COMPARISON OF MARRIED AND UNMARRIED FEMALES ON SELF-DISCLOSURE TO MOTHER, FATHER, AND FEMALE FRIEND
(Curve for disclosure to Spouse is shown in each panel)

upon the spouse, and became more reticent toward other persons. In this sense, self-disclosure enters into relations similar to those of libido in psychoanalytic theory.

Parent-Cathexis and Self-Disclosure to Parents

Total Mother-Cathexis scores for the group of 31 nursing students correlated −.63 with scores for self-disclosure to Mother; Father-Cathexis scores correlated −.53 with scores for self-disclosure to Father. Both coefficients were significant beyond the .01 level for $df = 30$. The direction of the obtained values signifies that high self-disclosure was associated with positive feelings toward the parents, while low self-disclosure was associated with attitudes of dislike toward the parents.

DISCUSSION

These preliminary findings demonstrate that self-disclosure is measurable, and that the present method for assessing it has some validity. The questions now open for exploration are virtually without limit, in view of the many possible and relevant combinations of the main factors—groups, target-persons, aspects of self, and individual differences.

Some questions suggested by the present findings may serve as guides to further exploration: Why do Negro *S*s consistently disclose less about themselves than whites, and why do females disclose more than males? Why is the mother the preferred target of self-disclosure for this age group? What is the significance of the fact that some aspects of self, for instance, Tastes and Interests, Attitudes and Opinions, and Work, are disclosed more than information about Personality, Money, and Body? Is it an artifact of the questionnaire, or does it reflect cultural consensus about what is readily disclosable and what is not? What individual traits besides feelings and attitudes toward target-persons account for individual differences in self-disclosure?

SUMMARY

A reliable questionnaire for the assessment of self-disclosure was described. Groups of both sexes, white and Negro, married and unmarried, were tested with the questionnaire for extent of self-disclosure of six different aspects of self to various target-persons—Mother, Father, Male Friend, Female Friend, and/or Spouse. The findings are summarized as follows:

1. Young unmarried *S*s, both white and Negro, showed the highest self-disclosure to Mother, with lesser amounts to Father, Male Friend, and Female Friend.

2. *S*s tended to vary the amount of self-disclosure with respect to the category of information to which an item about the self belonged. Two clusters of aspects emerged, a high disclosure cluster including Attitudes and Opinions, Tastes and Interests, and Work, and a low disclosure cluster comprised of Money, Personality, and Body.

3. White *S*s disclosed more than Negroes, and females more than males.

4. There was significant interaction among groups of *S*s, target-persons, and aspects of self.

5. Married *S*s disclosed less to Mother, Father, and Same-sex Friend than comparable unmarried *S*s. The married *S*s disclosed more to their Spouses than to any other target-persons. There was more disclosure to Spouses than to any other target-person on the part of the married and the unmarried groups.

6. A significant correlation was found between parent-cathexis and self-disclosure to the parents. The more that the parents were liked, the more disclosures were made to them.

REFERENCES

1. BLOCK, J. The assessment of communication. Role variations as a function of interactional context. *J. Pers.*, 1952, **21**, 272–286.
2. BLOCK, J., & BENNETT, LILLIAN. The assessment of communication. Perception and transmission as a function of the social situation. *Hum. Relat.*, 1955, **8**, 317–325.
3. FROMM, E. *Man for himself*. New York: Rinehart, 1947.
4. HORNEY, KAREN. *Neurosis and human growth*. New York: Norton, 1950.
5. JOURARD, S. M. Identification, parent-cathexis, and self-esteem. *J. consult. Psychol.*, 1957, **21**, 375.
6. JOURARD, S. M. *Personal adjustment: An approach through the study of healthy personality*. New York: Macmillan, 1958, in press.
7. LEWIN, K. Some social-psychological differences between the United States and Germany. In G. Lewin (Ed.), *Resolving social conflicts: Selected papers on group dynamics, 1935–1946*. New York: Harper, 1948.
8. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
9. RIESMAN, D. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.

Received April 15, 1957.

TWO FORMS OF SOCIAL CONFORMITY: ACQUIESCENCE AND CONVENTIONALITY[1]

HALLA BELOFF

Queen's University, Belfast

In current psychological literature, the term conformity is applied to two distinguishable phenomena, which may be termed acquiescence and conventionality.

Acquiescence refers to the agreement with expressed group opinion in a particular experimental situation involving pressure from others. In these experiments, it is commonly arranged that the personal social influence should give rise to inconsistency with previously elicited responses, or to a seeming distortion of perceptual judgment. Studies of such dynamic conformity have been common in recent years (1, 6, 7, 9, 10, 12, 14, 15, 17, 20, 21, 22). A number of experimental devices or ruses have been employed, all of which appear to elicit a significant degree of conformity with or shift towards group opinion. These devices would seem to vary principally in the degree to which they do or do not insult the intelligence of the subjects, i.e., in the subtlety with which the pressure is applied.

By *conventionality* is meant, of course, the concurrence with the tenets, attitudes, and mores of a subject's culture or subculture. This is a very broad concept referring to the characteristics of the "insider," as he may be opposed to the individualist or eccentric. Empirical studies are fewer here, but Brodbeck *et al.* (8) have submitted the Riesman conformity typology to objective test and Bernberg (4) has related conventionality in social perception or opinion to other personality traits and to objective life criteria.

The present study was designed to obtain some objective measure of both kinds of conformity phenomena and to investigate their relationship to each other. Conventionality is here viewed as the summated end-product of past specific or piecemeal acquiescences with cultural norms. The general hypothesis, then, is that those subjects who acquiesce readily in the experimental situation are also relatively conventional; and conversely, that subjects who are independent or intransigent are also relatively unconventional. To what extent, in brief, is the use of the word conformity for both concepts justified by empirical evidence? Further relationships to three pertinent personality traits are also investigated.

[1] This study was carried out while the author held a personal grant from the Medical Research Council, London.

METHOD

The Acquiescence Experiment

The experimental procedure followed that of Helson *et al.* (13), using the Blake simulated group technique (6).[2] Readers are referred to the original paper for a full description.

Subjects (*Ss*)[3] completed a shortened version of the Thurstone-Chave War Scale (paper-and-pencil) in the usual individual way. After an interval of approximately three weeks they again responded to the Thurstone-Chave statements, under simulated group conditions; that is, they heard the stimulus statements, plus four pseudo-respondents[4] pre-recorded on sound tape, with time left between items for their spoken responses as fifth *S*. The responses of the naive *Ss* were actually relayed to an adjoining room and recorded there.

The pseudo-*Ss* presented strongly agreeing, neutral and strongly disagreeing backgrounds in random order, and balanced with regard to anti-, pro-, and neutral-towards-war statements.

The present situation differed from that used by Blake in one important particular. It was not practically feasible to persuade the *Ss* that four others were present in other rooms concurrently. Instead, each *S* was told that the project was investigating aural attitude measurement devices and that he happened to be the fifth person to be recorded on this tape.[5] At the end of the testing period the nature and

[2] It presents a number of advantages over other systems: (*a*) it eliminates the need for the repeated presence of paid participants, (*b*) conditions within the procedure are constant from *S* to *S*, (*c*) no specially constructed apparatus is necessary. Further, the group information is presented obliquely, in contrast to those devices in which information about group decisions is presented as such, and the *Ss* are invited directly to give their own responses in the light of it.

[3] A mixed class of undergraduate students attending an elementary Psychology course. For the final analysis, 35 men and 25 women, who had completed all parts of the programme, were used.

[4] These four who provided the "other *S*" stimuli were advanced students, unknown to the naive *Ss*.

[5] One *S* only admitted suspecting a hoax, deduced from the fact that the other supposed *Ss*, while agreeing with each other, presented such logically incon-

purpose of the whole experiment were fully explained to the class.

Acquiescence scores. A measure of acquiescence was computed by obtaining the algebraic sum of response shift on the seven-point answer scale, towards and away from the background reports. That is, the difference between the first and second response sets was considered with respect to the group report, a shift towards the background being designated positive and a shift away negative. For each *S*, the shift scores were summated for all items in which the background report had been strongly agree or strongly disagree, 12 items in all.

The range of acquiescence, or net-shift scores, was found to be −6 to +37, with a mean and mode at +11 steps (C.R. of deviation from zero = 9.18). Blake has not given a reliability coefficient in any of his reports. The present data give a Spearman-Brown corrected consistency coefficient of +0.63.

Conventionality Measures

For the purpose of empirical study, conventionality was operationally defined as the degree of agreement (expressed by a rank correlation, for example) between an individual *S* and the mean or modal response of the rest of the group. Two measures were obtained using widely disparate fields of socialized behaviour, aesthetics, and politics.

Aesthetic conventionality. The data here were the results of a simple experiment in aesthetic preference. The stimulus materials were black and white photographic slides of eleven ceramic teapots in the collection of the Victoria and Albert Museum, London. The pots ranged in time from the 17th century to the present day and included oriental as well as European products. The test was carried out as a group experiment, using the method of paired comparisons.

A rank order of preference was computed for the pots, for each *S* from his comparisons, and a mean rank order for the whole group from these. A measure of the reliability of the preferences expressed is given by the mean *zeta*, Kendall's coefficient of consistence (16, p. 123), of 0.90 for the class.

A measure of aesthetic preference or taste having been obtained, aesthetic conventionality was then measured for each *S* by the Spearman rank correlation between individual rank order and the mean rank order of the whole group (no correction being made for the inclusion of the individual's order in the group). The range of these rank correlations was found to be −.095 to +.918 (i.e., no *S* showed real opposition to class taste), with a mean at +.633.

Politicosocial conventionality. The measure again refers to the degree to which *S*s conform to the popular response of their subculture—here with regard to statements concerning a variety of religious, social, and political attitudes. Eysenck's Inventory of Social Attitudes was used as a stimulus basis. The responses to individual items provided the data for the computation of Floyd Allport's Index of Institutionalization (5; 18, pp. 82–84).

With the arrangement of a five-point multiple answer choice, a value of 2.00 for this statistic for a particular item denotes the null case, i.e., an equal distribution of checks in each category. A value of 3.00 denotes a normal distribution, i.e., some social institutionalization and conformity to it; further increases in the value of the index occur as the extent of conformity increases.

Eighteen of the 40 items manifested institutional indices above 3.00 (range 3.01 to 7.81, mean = 4.46). These items were used to compute a weighted index of political conventionality in the following manner. For each item, the weighted score consisted of the Institutional Index multiplied by the number of steps the individual's check is removed from the modal category. The item scores were summated and would in this form give an *unconventionality* score. For the sake of uniformity, this was subtracted from the maximum possible score to give a score for conventionality. The theoretically possible range is zero to 286.9, the actual scores ranging from 160.1 to 272.0, with a mean at 226.4.

RESULTS

Acquiescence and Conventionality

A simple preliminary analysis was carried out to test the hypothesis that the conformity scores of the women would be higher than those of the men (as suggested by Crutchfield, 10). Table 1 shows that this is so for acquiescence (although the difference is not significant), and very significantly so for politicosocial conventionality. In view of these sex differences, the main analysis was carried out separately for men and women. Because at least one of the variables considered does not lend itself to the usual parametric analysis (aesthetic conventionality with a markedly negatively skewed distribution), a high power nonparametric was used, the Mann-Whitney *U* test (19).

Inspection of Fig. 1 and Table 2 shows that in all cases the relationships are in the pre-

TABLE 1

SEX DIFFERENCES IN DEGREE OF CONFORMITY (Mann-Whitney *U* Test between 35 men and 25 women on the basis of a single rank order)*

Variable	Men Mean Rank	Women Mean Rank	*U*	*p*
Acquiescence	28.7	33.0	374	.17
Aesthetic conventionality	30.6	30.5	—	—
Political conventionality	22.8	36.0	245	.002

* One-tailed, as predictions were made.

Note.—For each variable in this study Rank 1 was assigned to the *lowest* score.

sistent responses to the attitude statements. (He was, of course, discarded from the sample.)

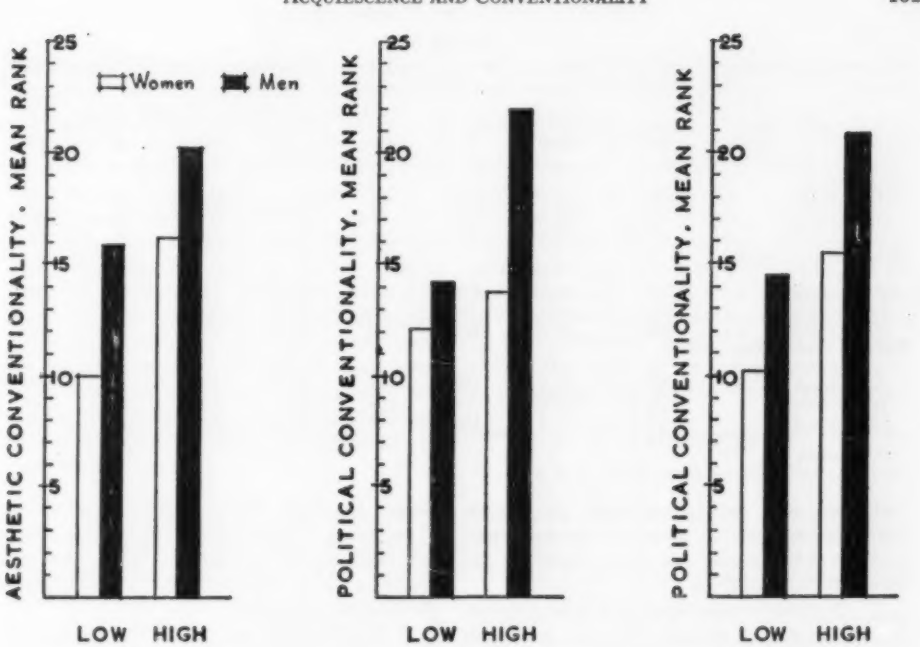

FIG. 1. RELATIONSHIP BETWEEN THE THREE CONFORMITY MEASURES
(high–low dichotomy at the median)

Note.—For the significance of the differences, see Table 2. As the number of men and women *S*'s are not identical, mean ranks as in this figure and Table 3 are meaningfully comparable only within each sex.

dicted direction. It is seen that conventionality, in the two fields, shows the most consistent connection, reaching significance at the .05 level for both men and women.

The hypothesis concerning acquiescence and its relation to conventionality is confirmed, although the sex differences that appear complicate the conclusion. As is seen, for men the acquiescence-politics relationship is significant, and for women the relationship between acquiescence and aesthetic conventionality. It is tempting to suggest that the former field is one that is more salient for men in our culture, and the latter for women; and that the salient attitudes only, partake in the orderly integration of the *socius*. Further research, using perhaps especially selected groups of designers and students of politics would elucidate this point.

Broadly it may be concluded that acquiescence and conventionality do go together; and that this study has presented further evidence therefore for a generalized tendency towards and away from conformity.

TABLE 2

RELATIONSHIP BETWEEN THE THREE MEASURES OF CONFORMITY

Variables	Men		Women	
	U	*p**	*U*	*p**
Acquiescence & aesthetic conventionality	115	.11	39	.02
Acquiescence & political conventionality	85	.02	67	.27
Aesthetic & political conventionality	96	.03	45	.05

* One-tailed.

Personality Test Correlates

The three personality traits considered here were chosen both for their seemingly obvious

TABLE 3

RELATIONSHIP BETWEEN CONFORMITY MEASURES (MEDIAN SPLITS) AND PERSONALITY VARIABLES (MEAN RANKS)

Personality Trait	Men				Women			
	Low	High	*U*	*p**	Low	High	*U*	*p**
Acquiescence								
Ascendance	20.1	15.7	114	.10	11.7	14.4	61	.18
Neuroticism	15.0	20.8	103	.05	15.7	10.0	42	.03
Calif. F score	15.4	20.5	108	.07	11.0	15.1	52	.08
Aesthetic Conventionality								
Ascendance	18.8	17.1	137	.31	11.3	14.8	56	.12
Neuroticism	21.1	14.3	93	.03	15.2	10.6	49	.06
Calif. F score	15.7	20.6	110	.08	12.5	13.5	72	.37
Political Conventionality								
Ascendance	18.3	17.7	149	.42	10.7	15.4	50	.06
Neuroticism	21.0	14.8	99	.04	13.4	12.6	73	.39
Calif. F score	15.6	20.2	113	.09	12.3	13.7	70	.32

* One-tailed.

relevance, and because although they have been the subject of previous investigation, findings with regard to them have by no means been consistent.

1. *Ascendance-submission.* It has been shown (13) that male *S*s who acquiesce in a simulated group situation have significantly lower ascendance scores on the A-S Reaction Study inventory. This is hardly a surprising finding, in that any of the acquiescence devices might well have been designed just as an objective validating experiment for the inventory.

From Table 3 it can be seen that the male *S*s here do display a tendency in this direction; higher scores on all three forms of conformity going with lower ascendance scores. The women, however, manifest a consistent trend in the opposite direction, the more conforming *S*s being the *more* ascendant ones.

2. *Neuroticism.* The consensus of opinion with regard to neuroticism seems to be that although independent nonconformists admit to having more neurotic tendencies when questioned more or less directly (2, 9, 15), it is the conformers who are perhaps *really* more neurotic (10, 14). Using the Maudsley Personality Inventory designed by C. M. Fransk (11), it was predicted that this simple and rather obvious self-rating procedure would show the conformers to be less neurotic.

Table 2 shows that again the results are by no means simple. For conventionality, those data that reach significance or approach it, do confirm the prediction, i.e., the unconventional *S*s show higher neuroticism scores. Considering acquiescence however, the women too confirm the hypothesis, while the men attain a result of equivalent significance in the opposed direction.

3. *Authoritarianism.* One of the assumptions both overt and covert in the Californian writings on the authoritarian personality is that authoritarian persons are conformers. And indeed this is the one consistent relationship that has been demonstrated in all the studies concerned with it. Some report correlations between acquiescence and actual scores on the California E or F scales (7, 10, 20): others with ratings, etc., of traits underlying the authoritarian concept (2, 14).

Using the F scale (Forms 45 and 40), that connection is confirmed here, albeit only at the .07 and .08 confidence levels for men and women respectively. With regard to conventionality only the men's results approach significance.

However, before accepting a simple link between the tendency towards authoritarianism and conformity, as has been done so far by other writers, it is necessary to consider the school of thought (3) which has considered that a large part of the variance of the F scale in its present form is taken up by some behaviour which is best called acquiescence! In the light of the evidence at present available, the most economic explanation of the acquiescence-F score relationship found here and in previous studies must be that the

acquiescence-inducing situations and the orthodox F scale just measure the same thing.

Obviously, further investigation must now concern itself not with the development of yet more acquiescence devices but with the clarification of the place of conformity in general personality organization and, more particular, with the meaning and significance that conformity has for men and women, in order to elucidate the part that conformity plays in their adjustment to their respective roles. This meaning or significance would appear to be very different for the two sexes. As we have seen in our data, acquiescent men appear to be more submissive and more neurotic, acquiescent women more ascendant and more stable.

SUMMARY AND CONCLUSION

In this project, the two principal aspects of conformity were empirically studied; i.e., the dynamic process of immediate conformity, termed acquiescence; and long term results of conformity to cultural norms, termed conventionality.

The relationship between the two phenomena was studied, as were correlates with three seemingly pertinent personality characteristics. The main hypothesis put forward was that persons scoring relatively high on the acquiescence measure would tend also to score high on conventionality indices.

On the basis of performance in a Blake-type simulated group situation, a group of 60 undergraduate university students were divided into two categories of high and low acquiescence. Conventionality, operationally defined as high agreement between an individual's response and the mean or modal response of his group, was measured in the areas of aesthetics and politics. Photographs of teapots and the Eysenck Social Attitudes Inventory provided stimulus material for the computation of conventionality scores by means of rank correlations and the Allport Institutional Index.

Analysis by means of the Mann-Whitney *U* test confirmed the hypothesis of a positive relationship between all three conformity scores. Certain sex differences appeared: women *S*s tended to exhibit a greater degree of conformity than men; and also, the acquiescence-conventionality relationship reached significance for men in politics and for women in aesthetics.

The literature had indicated that present knowledge of conformity-personality correlates is in a confused state. The present data confirm this point. Striking sex differences again add complexity to the results. The only relationship that is found to be uniform for both sexes and is congruent with past reports is that between acquiescence and high scores on the California F Scale. The possibility that this is the result of an experimental artifact in the present form of the F scale is discussed.

REFERENCES

1. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
2. BARRON, F. Some personality correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287–297.
3. BASS, B. M. Authoritarianism and acquiescence. *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
4. BERNBERG, R. E. Personality correlates of social conformity. *J. appl. Psychol.*, 1954, **38**, 148–149.
5. BERNBERG, R. E. A measure of social conformity. *J. Psychol.*, 1955, **39**, 89–96.
6. BLAKE, R. R., & BREHM, J. W. The use of tape recording to simulate a group atmosphere. *J. abnorm. soc. Psychol.*, 1954, **49**, 311–313.
7. BLOCK, JEANNE, & BLOCK, J. An interpersonal experiment on reactions to authority. *Hum. Relat.*, 1952, **5**, 91–98.
8. BRODBECK, A. J., NOGEE, P., & DI MASCIO, A. Two kinds of conformity: A study of the Riesman typology applied to standards of parental discipline. *J. Psychol.*, 1956, **41**, 23–45.
9. CERVIN, V. Experimental investigation of behaviour in social situations: I. Behaviour under opposition. *Canad. J. Psychol.*, 1955, **9**, 107–116.
10. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
11. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, in press.
12. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
13. HELSON, H., BLAKE, R. R., MOUTON, JANE S., & OLMSTEAD, J. A. Attitudes as adjustments to stimulus, background and residual factors. *J. abnorm. soc. Psychol.*, 1956, **52**, 314–322.
14. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
15. JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663–667.
16. KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1948.

17. OLMSTEAD, J. A., & BLAKE, R. R. The use of simulated groups to produce modifications in judgment. *J. Pers.*, 1955, **23**, 335–345.
18. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
19. SIEGEL, S. *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
20. WELLS, W. D., WEINERT, G., & RUBEL, MARILYN. Conformity pressure and authoritarian personality. *J. Psychol.*, 1956, **42**, 133–136.
21. WIENER, M., CARPENTER, JANETH T., & CARPENTER, B. External validation of a measure of conformity behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 421–422.
22. YOUNG, N., & GAIER, E. L. A preliminary investigation into the prediction of suggestibility from selected personality variables. *J. soc. Psychol.*, 1953, **37**, 53–60.

Received April 17, 1957.

THE ROLE OF MOTIVATION IN PSYCHOLOGICAL STRESS[1]

WILLIAM VOGEL, ROBERT W. BAKER, AND RICHARD S. LAZARUS[2]

Clark University

LAZARUS and Baker (5) recently have presented a theoretical and methodological framework for the study of behavior under stress. Their conception is that the motivational state of the individual in relation to the stressor conditions imposed upon him determines whether or not a state of stress occurs. Experimentation within this framework thus involves the systematic use of motivation as an independent variable. The purpose of this paper is to present some evidence concerning the role of motivation in determining response to stressor conditions.

In our experimental studies we distinguish between *induced* and *intrinsic* motivation. By induced motivation we mean those transitory organismic states which we expect to arouse or instil through the use of ego-involving instructions or through the manipulation of incentives. Intrinsic motivation, on the other hand, refers to motive states that are presumed to be relatively enduring characteristics of individuals. These are typically inferred from apposite behavioral samples, including fantasy productions (cf. McClelland, 7).

A further distinction may be made between kinds of behavior samples from which intrinsic motivational states are inferred. On the one hand, "long run, everyday" behavioral characteristics of individuals may be used, as when achievement motivation is inferred from the number of hours that a student studies per week. On the other hand, imaginational production or fantasy may be employed, as McClelland and his associates have done in the case of achievement (7), affiliation (1, 10), and power motivation (13). The assumption is often made that the incidence of need related imagery in the stories a person tells to picture stimuli (especially in the absence of external achievement cues) reflects a relatively stable and persistent motivational characteristic.

All these means of studying motivation are relevant to the three theoretical issues to which the present experiment is addressed. The first of these issues involves the relationship between a stressor condition[3] and motivational characteristics of individuals. If there is no relevant motive state to be thwarted, then according to our theoretical position there is no stress; if there is a strong motive state, but it is not engaged by the stressor condition, then there is likewise no stress; if there is a motive state, and it is engaged by the stressor condition, the resultant stress state and its consequences vary as a function of the intensity of the motive state. In this study, we focus upon the last of these alternatives and investigate behavioral output and affective arousal as a function of induced motive states varying in intensity. If differential achievement motivation is induced in two groups of subjects, we would expect that failure-information (the stressor condition) should be more thwarting to the more highly motivated subjects, resulting in greater affective arousal (i.e., heightened physiological reactivity) and a greater increase in effort on the task.

The second problem addressed by our experiment is the generality of motive states, i.e., the extent to which a motivational characteristic inferred from one behavioral context is also operative in another. By definition, this problem refers to intrinsic rather than induced motivation. The study seeks to determine whether degree of motivation inferred from level of academic effort and accomplishment in college is associated with behavioral output on a performance task and accompanying physiological reactivity under various experimental conditions, including a stressor. The experimental design also permits the effect of intrinsic motivation on behavioral output and physiological reactivity to be compared with those of induced motivation.

The third issue to be considered concerns the relationship between the expression of a motive in imaginational productions or fantasy

[1] This investigation was supported by research grant (M-734) from the National Institute of Mental Health, Public Health Service.

[2] Now at the University of California at Berkeley.

[3] By stressor conditions we mean those experimental conditions which are intended by the investigator to be stressful, but which are not necessarily responded to as such by the *S*. The term "stress" we use to refer to a state of the organism, not to a class of conditions.

and its actualization in behavior. Fantasy has been used as a means of measuring intrinsic motivation on the assumption that this relationship is always direct and positive (7), but evidence has recently been presented that this assumption is not entirely warranted (6). The present experiment investigates the relation of a fantasy measure of need for achievement (7) to behavioral output and physiological reactivity under various experimental conditions, and also compares this measure with indices derived from other behavioral sources.

METHOD

Subjects and Procedure

Forty male undergraduate college students (mostly freshmen) who volunteered for the study were selectively assigned to two groups of 20 each in such a manner that the two groups represented similar distributions of scores on two measures of intrinsic motivation, McClelland's need Achievement score and a measure relating academic accomplishment to intellectual ability. One group was told that they were going to perform a task that would allow the experimenter to evaluate the effect of work on physiological activity. This information was intended to be relatively less motivating or ego-involving than the treatment given the second group, which was informed that their performance on the task would provide the experimenter with a means of assessing their potential for high level academic and vocational functioning. All of the *S*s in the experiment were tested individually.

Before the differentially motivating instructions were given, the two groups were given a "practice" session on the experimental task, ostensibly in order to become familiar with it. Actually, this session ensured the comparability of the two groups regarding physiological reactivity, as well as affording a baseline for gauging the effects of conditions. The practice condition lasted approximately four minutes during which all *S*s performed an equal number of task items.

The experimental task was the McKinney Reporting Test (8), which presents a long series of circles containing three symbols in various quantities. The *S*'s task is to count as quickly and accurately as possible the number of symbols of each kind which each circle contains.

Following the practice period, the differentially motivating instructions were given and a 12-minute "work" condition ensued. At the end of this period, *S*s were informed that their performance was inadequate and that they would be given an opportunity to do better. During a second ensuing 12-minute period of performance (the failure-stressor condition) both groups of *S*s were given false, unattainable norms every minute to induce a sense of failure.

Following the failure-stressor condition, *S*s were administered cards 6BM and 7BM of the Thematic Apperception Test and the machineshop picture of the McClelland series (7), which they had taken under neutral conditions on an earlier day. If a different set had been used the second time following McClelland's procedure, it would not have been clear whether obtained differences in imagery were a function of conditions or of differential achievement cues of the cards. McClelland's attempt to equate the cue value of pictures is of questionable value in studies of individual differences, because the cue value of a picture may vary for different individuals and *S* populations.

At the same earlier session, the Taylor Anxiety scale (12) was also administered, and *S*s were asked to indicate on a chart their activities for the previous week hour by hour to provide reliable information about their study habits.

Physiological Measures

Measures of physiological reactivity were taken as follows. In the practice condition, galvanic skin response (GSR) readings were made during the first and third minutes, pulse was recorded at two minutes, and diastolic and systolic blood pressure at four minutes. The same procedure was followed in the two subsequent conditions, being repeated three times during each 12-minute period.

The GSR apparatus used was a type 310-A Z-Angle meter, manufactured by The Technology Instrument Corporation of Acton, Mass. An internal oscillator was used, operating with 400 cps sources of test signal. Approximately .04 volts of AC current was impressed across the unknown impedance. Two electrodes made from dimes were placed, after electrode paste had been applied, to the first and index fingers of the *S*'s left hand. Direct impedance measurements were recorded.

The GSR reading employed was the average of three readings taken at 20-second intervals over a one-minute span. The pulse recorded was based on a constant reading taken over a one-minute span. Blood pressure was taken by the usual auscultatory method with a standard mercury sphygmomanometer. Diastolic blood pressure was read at the point of disappearance of the sound. For both diastolic and systolic blood pressure, two readings were taken per minute, and the average of the two was recorded as the reading for that minute. Of the three diastolic pressures, three systolic pressures, and three pulse readings obtained for each condition, the peak response for each measure was retained for further analysis. Of the six GSR responses recorded for each condition, the lowest impedance reading was retained for further analysis. All impedance readings were converted into conductance scores.

The index of physiological reactivity employed was adapted from Lacey (3). The measures of systolic and diastolic blood pressure, pulse, and GSR were separately normalized and converted into *T* scores. The most reactive measure for each *S* across the several conditions was then determined by adding *T* scores across conditions. The channel for each individual having the highest *T*-score total was selected as his most reactive channel, and his index of physiological reactivity was based on data from this channel. This index is expressed as a logarithmic transformation of percentage change from the practice condition to the work and stressor conditions respectively, thus correcting for base level.

Lacey's method has its drawbacks, as the most

reactive measure over the course of the experiment might not be most reactive in a particular condition. However, 66 per cent of the *Ss* showed perfect consistency through all conditions, and another 21 per cent were highly consistent in that measurements in their most reactive channel did not differ by more than 2 *T*-score units from their most reactive measure in any single condition.

There are essentially two alternatives to Lacey's technique. One is an index based on an average of scores from all autonomic channels. This method involves the questionable assumption that a person's somatic reactivity is a direct function of the number of channels in which he expresses this reactivity, although it avoids the assumption that the highest physiological response is the best measure of arousal. The authors obtained very few significant results whenever it was applied.

The other alternative is to utilize the *S*'s most reactive measure in *each* condition. This method, which is consistent with an assumption of extreme specificity, has the drawback of giving maximal weight to chance autonomic fluctuations. A larger number of uninterpretable results were derived with this method than with any other. Accordingly, Lacey's method seemed the best for the present study. Many of these problems in the use of autonomic measures in psychological research are discussed by Lacey in a recent article (4).

RESULTS AND DISCUSSION

Relationship Between a Stressor Condition and Motive States

Table 1 shows change in physiological reactivity from the baseline measures of the "practice" condition under the "work" condition, after the different instructions had been given the *Ss*; and under the stressor condition, when both groups of differentially motivated *Ss* were given failure-information.

It will be noticed that in the work condition, the more strongly motivated of the two groups shows the greater increase in physiological reactivity. However, and this we did not anticipate, physiological reactivity in the two groups is at the same level under the stressor condition.

Table 1 also shows that similar results are obtained with performance output on the McKinney Reporting Test, i.e., number of items attempted.[4] In the work condition, the strongly motivated group attempts significantly more items than the weakly motivated group, while in the stressor condition both groups increase in output, the *greater* increase in the weakly motivated group rendering the two groups again comparable.

TABLE 1

PHYSIOLOGICAL REACTIVITY AND TASK PERFORMANCE AS FUNCTIONS OF INDUCED MOTIVATION AND EXPERIMENTAL CONDITIONS

Groups and Conditions	*T* Scores of Logarithm Percent Physiological Change*	Items Attempted McKinney Test**
High induced motivation ($N = 20$)		
Work	60.42	35.04
Failure-stressor	65.43	40.89
Low induced motivation ($N = 20$)		
Work	52.29	27.89
Failure-stressor	60.82	40.09

* Interaction significant below the .05 level ($F = 4.58$).
** Interaction significant below the .01 level ($F = 8.98$).

The findings presented in Table 1 do not support our hypothesis that different levels of induced motivation are associated with differential response to a stressor condition. Our first inclination was to interpret these results as suggesting that for the relatively unmotivated *Ss*, failure-information had an effect equivalent to that of the motivating instructions for the other group, making level of physiological reactivity and performance output comparable for the two groups. Our second inclination was to look for interactive effects which, as we had learned in previous studies (5), could obscure real differences. Accordingly, we examined the relationship between level of induced motivation, performance output, and physiological reactivity under the stressor condition.

We first determined that there is no relationship between physiological reactivity and performance output under the stressor condition when induced motive state is not taken into account. That is, *Ss* with marked physiological reactivity are just as likely to have low as high behavioral output. However, when these two variables are considered in their relationship to level of induced motivation, a strong interaction appears.

Table 2 shows number of items attempted on the McKinney Reporting Test under the stressor condition as a function of level of induced motivation and degree of physiological reactivity. It is clear from this table that strongly motivated *Ss* who were most reactive physiologically attempted signifi-

[4] Errors do not account for an appreciable amount of variation and are therefore not presented as a dependent variable; number of items attempted and number of items correct correlate very highly.

TABLE 2
TASK PERFORMANCE UNDER FAILURE AND MAS SCORE AS FUNCTIONS OF INDUCED MOTIVATION AND LEVEL OF PHYSIOLOGICAL REACTIVITY
(N = 6 in each group)

Groups	Items Attempted, McKinney Test*	MAS Score**
High induced motivation		
Somatically reactive	43.39	10.0
Somatically nonreactive	34.00	14.5
Low induced motivation		
Somatically reactive	35.44	16.2
Somatically nonreactive	42.72	12.5

* Interaction significant below the .01 level (F = 14.24).
** Interaction nonsignificant (F = 2.60; that required for significance at the .05 level = 4.35).

cantly more items than strongly motivated *S*s who were least reactive. On the other hand, *weakly* motivated *S*s who were most reactive physiologically attempted significantly *fewer* items than weakly motivated *S*s who were least reactive. In other words, under strong motivation the relationship between physiological reactivity and performance output is *positive*, while under weak motivation it is *negative*. These findings with respect to induced motivation had no parallel with respect to intrinsic motivation.

Why should physiological reactivity be positively related to performance under strong induced motivation, and negatively under weak motivation? It seems appropriate that highly motivated *S*s, told that performance was very important, should react to failure-information with relatively greater physiological response and that less motivated *S*s, given to believe that performance was of no great consequence, should react with relatively little physiological response. By the same token, it seems *in*appropriate that highly motivated *S*s should show relatively little physiological response to failure-information, or that weakly motivated *S*s should show relatively greater reactivity.

Interestingly, those individuals responding inappropriately—highly motivated-physiologically unreactive; weakly motivated-physiologically reactive—show less performance output on the experimental task than those who respond appropriately under the stressor condition. Possibly the inappropriate responders, being also less productive, may be the less well-adjusted *S*s. Since the Taylor Manifest Anxiety Scale (MAS) has been demonstrated to correlate highly with a scale of neuroticism (14), this hypothesis can be checked; the relevant data are in the second column of Table 2. The differences are not statistically significant, but there is a suggestive tendency for the inappropriate responders to have higher scores.

The Generality of an Intrinsic Motive State

The availability for our *S*s of academic records and a measure of intellectual ability made it possible to devise a motivational measure in terms of the relationship between accomplishment and capacity. "Overachievers" and "underachievers" were identified by subtracting each student's score (expressed in standard score form) on the American Council on Education Psychological Examination for College Freshmen (administered in all cases just prior to his matriculation as a freshman) from his college grade point average (also expressed in standard score form).

We assumed that overachievers tend to be highly motivated to achieve and underachievers poorly motivated. In present terminology, the difference, if it be motivational, clearly pertains to intrinsic motivation. The motivational interpretation of over- and underachievement is substantiated by the measure of study habits. Overachievers reported an average of 34 hours spent studying for the week, while underachievers reported an average of 14 hours.

We used this index of intrinsic motivation as an independent variable in the prediction of behavior in the experimental situation, expecting that overachievers would attempt more items and show more accompanying physiological activity. The data for this comparison are presented in Table 3. It is clear that our expectations were not realized. The underachievers have greater performance output and are more physiologically reactive than the overachievers.

We would account for this apparent paradox by suggesting that we attributed unwarranted generality to the intrinsic motive state, neglecting the important interactive relationship that may be presumed to exist between a motive state and the operations required by a task. We failed to take into account the

TABLE 3

NUMBER OF ITEMS ATTEMPTED ON THE MCKINNEY REPORTING TEST AND SOMATIC REACTIVITY AS A FUNCTION OF INTRINSIC MOTIVATION

Groups (N = 20 in each case)	Items Attempted, McKinney Test**	Somatic Reactivity as T Scores of Logarithm Percent Physiological Change*
Overachievers	31.98	58.07
Underachievers	39.02	62.46

Note.—Performance data for both groups have been combined across all conditions since there were no significant interactions with conditions; physiological data were combined across the work and stressor conditions only, since they are expressed as log percentage changes from the practice condition.

* Difference significant below the .05 level (F = 4.47).

** Difference significant below the .01 level (F = 10.39).

differences between academic tasks and the McKinney test, the former requiring conceptual operations, the latter being a simple, repetitive, monotonous, perceptual-motor task. After the fact, it seems plausible that the overachievers, being motivated to achieve in conceptual tasks, were "left cold" by the McKinney test; i.e., the test failed to engage the achievement motivation which they had demonstrated in another context, while it did engage the achievement motivation of the underachievers by virtue of its perceptual-motor character. Unpublished questionnaire observations by the authors have suggested that underachievers had greater interest in motoric tasks, such as athletics, while overachievers spent more time at, and expressed more interest in, academic and intellectual activity. This explanation remains to be tested. The implication that motive states should not be considered as independent of operations required by tasks, in view of individual differences in predilection for different operations, nevertheless now seems to us almost self-evident.

The Relationship Between Expression of a a Motive in Imaginational Productions and its Actualization in Behavior

In earlier studies (6) we used as a measure of intrinsic motivation, a method devised by McClelland and his associates (7) for inferring achievement motivation from imaginational productions to picture stimuli. However, results of such studies led us (6) to question two assumptions underlying the use of a fantasy index of motivation: (*a*) that need states are necessarily expressed in fantasy, and (*b*) that when they are, there is necessarily a direct and positive relationship between the amount of need-related imagery and the degree of actualization of the need in behavior. If a person shows striving for achievement in his everyday behavior, one might expect him also to express this need in his imaginational productions, and vice versa. If fantasy can also be conceived as serving as a substitutive activity in the absence of need gratification, however, the expression of need in fantasy need not correspond with its expression in other forms of behavior. Lazarus *et al.* (6) have advanced the view that whether a need is expressed in fantasy, in motoric activity, or both depends in large part upon regulative processes in the personality.

With the present data, the need Achievement score was examined in relation to performance output and physiological reactivity. If it is a "pure" index of achievement motivation, it might be expected to be positively associated with number of items attempted on the McKinney Reporting Test and with the index of physiological reactivity. Two sets of need Achievement scores are available for our *S*s: those derived from story-productions taken under innocuous conditions on the "control" day (hereafter referred to as "neutral"), and those derived on the "experimental" day, following an experience of failure (hereafter referred to as "aroused"). For purposes of analysis of variance, *S*s were ordered separately according to their neutral and their aroused need Achievement scores and divided in each case at the median.

No significant relationship with either dependent variable was found for the need Achievement measure taken under the neutral condition. Thus, when *S*s were administered the picture stimuli in the absence of external achievement cues, the amount of need-related imagery expressed does not predict behavior in the experimental context. In addition, there were no significant interactions between neutral or aroused need Achievement measures and the experimental conditions, including the stressor. Neither were there significant relationships between the need Achievement score and induced motivation: The highly motivated group did not have higher arousal scores than the weakly motivated group.

TABLE 4

NUMBER OF ITEMS ATTEMPTED ON THE MCKINNEY REPORTING TEST AS A FUNCTION OF AROUSED NEED ACHIEVEMENT

Groups (N = 20 in each case)	Items Attempted, McKinney Test*
High n Achievement	32.74
Low n Achievement	38.51

* Difference is significant below the .05 level (F = 5.71).

There was, however, a main effect for the aroused need Achievement measure, but it was in the direction opposite to what would have been predicted on the basis of an assumed direct and positive relationship between the fantasy and behavioral expression of needs in performance output. Table 4 shows that, independent of conditions, persons with relatively greater expression of achievement-related imagery after being given failure information attempt *fewer* items on the experimental task than persons with lesser expression of such imagery.

This finding is comparable to those cited in the preceding section, where the over- and underachievement index was found to be negatively related to performance and physiological reactivity. In other words, here again one might argue that the McKinney Reporting Test does not challenge the *S*s with high achievement fantasy scores. While these results together suggest the possibility of a positive relationship between the need Achievement score and the over- and underachievement index, neither analysis of variance nor correlational methods reveal any such association. In relation to our negative relationship between need Achievement and performance on the McKinney Reporting Test, Atkinson and Reitman (2) have found no relationship between need Achievement and performance on a perceptual-motor task (drawing X's in circles), and a positive relationship between need Achievement (under neutral conditions) and performance on an arithmetic task. We have suggested earlier that a perceptual-motor task was not involving for the high need Achievers, while a conceptual task might be. The Atkinson and Reitman finding adds credence to this possibility, and also suggests that the nature of the need Achievement-performance relationships depends upon experimental conditions.

TABLE 5

PHYSIOLOGICAL REACTIVITY AND MAS SCORE AS A FUNCTION OF NEUTRAL AND AROUSED NEED ACHIEVEMENT AND EXPERIMENTAL CONDITIONS (N = 10 in each cell)

Achievement Imagery	T Scores of Logarithm Percent Physiological Change**		MAS Score*
	Work	Failure-stressor	
High Imagery, neutral condition	56.04	65.12	21.53
High imagery, aroused condition	56.82	61.16	11.30
Low Imagery, neutral condition	56.12	59.99	12.40
Low imagery, aroused condition	60.01	69.30	13.20

* Double interaction significant below the .05 level (F = 4.87).
** Triple interaction significant below the .01 level (F = 8.77).

In Table 5, evidence for a triple interaction between neutral and aroused need Achievement imagery and experimental conditions, including the failure stressor, is presented. According to this table, the most physiologically reactive group of *S*s, and the one also showing the most change in physiological activity from the work to the failure stressor condition, is the group which has the lowest need Achievement scores under both the neutral and aroused conditions. This finding is comparable to one by Raphaelson (9) who found that *S*s with low need Achievement scores (under aroused conditions) showed greater physiological reactivity (GSR) in a failure situation than high need Achievers. Raphaelson interpreted his results in terms of "fear motivating avoidance" on the part of the low achievement group, while we propose that there is more "positive interest" in a perceptual-motor task on the part of the low need Achievement *S*s, and less involvement on the part of the high need Achievers. Either interpretation would account for the negative relationship between need Achievement score and physiological reactivity. However, the additional finding, presented here, that there is also an inverse relationship between need Achievement and performance output, seems to support the notion that low need Achievers are more ego-involved in perceptual-motor

tasks while high need Achievers have greater involvement in abstract-conceptual tasks.

As a final comparison, we analyzed the relationship between the need Achievement scores under the neutral and aroused conditions and the Taylor Anxiety scale (12), obtaining the significant interaction also shown in Table 5. *Ss* with high need Achievement scores under both neutral and aroused conditions are seen to have the highest anxiety scores. In conjunction with the negative relationships between need Achievement scores and physiological reactivity and performance output, this relationship between need Achievement and anxiety score is difficult to interpret in terms of drive level as might be suggested by Taylor and others (11). Rather, we are inclined to believe that *Ss* giving high need Achievement scores, even in the absence of clear, external achievement cues (the neutral condition), give more evidence of neurotic disturbance. The foregoing results strongly suggest that the fantasy index should not be regarded as a pure measure of motivation that is actualized in behavior.

SUMMARY AND CONCLUSIONS

This paper reports an experiment designed to investigate motivational characteristics of individuals as a source of individual differences in response to stressor conditions. Motive states were manipulated in two essential ways: (*a*) through induction or arousal by use of ego-involving instructions (induced motivation) and (*b*) through inference from fantasy productions and academic accomplishment as well as study habits (intrinsic motivation). Effects of such motive states were sought in two kinds of behavior under varied experimental conditions, including a failure-stressor. The two kinds of behavior examined were performance output on a perceptual-motor task (the McKinney Reporting Test) and accompanying physiological reactivity.

The study is addressed to three theoretical issues. The first of these refers to the relationship between a stressor condition and induced motive states which vary in intensity. Our expectation that there would be a simple relationship between strength of induced motive state and behavior under a stressor condition was not borne out. However, different motive states are found to be associated with differential response to a stressor *when motivation is examined in interaction with level of physiological reactivity*. An attempt was made to account for the specific interactive effects in terms of degree of appropriateness of the reaction of the subject to the conditions.

The second theoretical issue involves the generality of an intrinsic motivational state inferred from the relation between academic accomplishment and intellectual capacity. Our results indicate that strength of this motive state is inversely related to performance output and physiological reactivity in the experimental situation. We attempt to account for this seeming paradox by suggesting that a motivational characteristic inferred from one behavioral context (i.e., the academic situation) is not necessarily operative in a different context (i.e., the McKinney Reporting Test).

The third theoretical issue of relevance in this experiment pertains to the relationship between the expression of a motive in imaginational productions or fantasy and its actualization in behavior. A positive and direct relationship is often assumed. That the relationship is not always positive and direct is shown by our findings, where amount of achievement-related imagery is inversely related to motivated behavior. We suggest that whether a need is expressed in fantasy, in behavior, or both will be a function of regulative processes in the personality.

The findings of this experiment indicate the complexity of the role of motivation in determining response to stressor conditions and highlight the methodological difficulties involved in the systematic investigation of motivational variables. At the same time, the results indicate the importance of investigating motivational variables in stress experimentation.

REFERENCES

1. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
2. ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–367.
3. LACEY, J. I. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, **43**, 338–350.
4. LACEY, J. I. The evaluation of autonomic re-

sponses: toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123–164.

5. LAZARUS, R. S., & BAKER, R. W. Personality and psychological stress—A theoretical and methodological framework. *Psychol. Newsltr.*, N.Y.U., 1956, **8**, 21–32.
6. LAZARUS, R. S., BAKER, R. W., BROVERMAN, D. M., & MAYER, J. Personality and psychological stress. *J. Pers.*, in press.
7. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. MCKINNEY, F., STROTHER, G. B., HINES, R. R., & ALLEE, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
9. RAPHAELSON, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–19.
10. SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
11. SPENCE, K. W., & TAYLOR, JANET. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
12. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
13. VEROFF, J. Development and validation of a projective measure of power motivation. *J. abnorm. soc. Psychol.*, 1957, **54**, 1–8.
14. WINNE, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117–122.

Received April 22, 1957.

INTERPERSONAL PREFERENCE AND COGNITIVE ORGANIZATION[1]

NATHAN KOGAN AND RENATO TAGIURI

Harvard University

THE present paper describes an empirical test of certain aspects of Heider's theory of cognitive organization (2). In general terms, this theory states that the relation between interpersonal attitudes and social perception is largely determined by simple cognitive configurations. We shall deal exclusively with the case in which the elements of the cognitive unit are three persons—the perceiver, "p"; the other, "o"; and a third person, "q," and, in particular, with Heider's L-relations as represented by positive and negative feelings existing among members of small groups. These L-relations generate eight basic situations (cf. Table 1). Unit #2, for example, corresponds to the case in which: p likes o, p dislikes q, and p thinks that o dislikes q.

According to Heider, units characterized by positive valences between the three component entities or by one positive and two negative valences are "balanced." Units containing three negative, or one negative and two positive valences are "imbalanced." With the assumption that people seek and attempt to maintain balanced cognitive structures while they aim to avoid and alter imbalanced ones, we shall test Heider's hypotheses by examining the incidence of two balanced and two imbalanced situations (Units #1 and #4, #6 and #7, Table 1). This necessarily requires the selection of appropriate baselines against which the observed frequencies can be compared. The baseline implicit in Heider's conceptual position is that all existing cognitive configurations will be balanced and none will be imbalanced. This is, however, an extremely harsh model for an empirical test. Furthermore, it fails to allow for the possibility of balance and imbalance as a continuous as opposed to a dichotomous process. There are two other baselines that provide relevant comparisons with observed values. The first, based on a chance expectancy model, is defined by the degree of balance that would be generated in the network if subjects were responding as random automata. The second baseline, based upon objective reality, is the actual level of balance present in the social unit. If cognitive balance exceeds this level, it would demonstrate that the cognitive process is sufficiently strong to impose a degree of balance that "overshoots" actuality. Our basic prediction is that *balanced* configurations will occur in *excess* of chance expectancy and objective reality. Correspondingly, we predict that the frequency of *imbalanced* configurations will be *below* these two baselines.

METHOD

Subjects and Procedure

Five groups of naval enlisted personnel with 12, 12, 15, 16, and 22 men, respectively, were used as *S*s. In each case, the men had been working and living together for at least two months on board ship as members of special teams.

The *S*s were requested, *first*, to indicate the three crew members with whom they would most like to spend a 72-hour liberty, and, *second*, to identify the three choices made by each of the other members. In the case of the 22-man group, which had been exposed to a great many research procedures, it was possible without unduly disturbing the members to obtain data regarding the three persons with whom they would *least* like to go on liberty.

RESULTS

Consider first the *balanced* Unit #1 (cf. Table 1). If the perceiver, p, chooses two persons, o and q, then p may expect a total of two choices to exist between o and q, i.e., o's choice of q and q's choice of o. If, however, as is the case in our procedure, p chooses three other members of the group, o, q, and r, then p may attribute or expect that o choose q, q choose o, o choose r, r choose o, q choose r, and r choose q. That is, p may expect a total of six choices to exist among the three persons he chooses. If the data for all group members are combined, the total number of attributed choices expected would be six times the number of perceivers, that is, six times the number of members in the group (cf. "Total possible column for balanced units in Table 2).

These maximum values represent perfect balance. As we suggested above, this is an exceedingly harsh baseline against which to

[1] This study was conducted at the Laboratory of Social Relations, Harvard University, under ONR Contract No. N5ori-07670. This paper was read at the EPA meetings, April 1957.

TABLE 1
COGNITIVE UNITS GENERATED BY HEIDER'S L-RELATIONS

Unit number	p's feeling for o (pLo)	p's feeling for q (pLq)	p's conception of o's feeling for q (oLq)	
1[a]	+	+	+	balanced
2	+	−	−	
3	−	+	−	
4[a]	−	−	+	
5[a]	+	+	−	imbalanced
6	+	−	+	
7	−	+	+	
8[a]	−	−	−	

[a] Units where pLo and pLq are of the same sign are transitive. Consequently, the unit includes both the oLq and qLo relationships.

both balanced units under consideration and are shown for each group in Column A, Table 2. A chi-square comparison of the total number of observed attributions with the number expected by chance in each of the five groups yields five ratios, all statistically significant beyond the .001 level (cf. A vs. B difference for balanced Unit #1 in Table 2). Clearly, the number of balanced configurations perceived by p deviates significantly from a chance baseline in the direction predicted by Heider's hypothesis.

Consider, next, the other baseline proposed above. Is the cognitive strain toward balance in excess of the corresponding relations existing in reality—i.e., the actual number of choices exchanged by the three chosen members? This baseline is quite stringent since the extent to which an *S*'s three favorites choose each other in fact already exceeds (significantly for the three larger groups) the chance baseline discussed above (cf. A vs. C difference for balanced Unit #1 in Table 2).[4] Comparison of attributed and actual choices shows, however, that cognitive balance exceeds objective balance in every case, and significantly so in four of the five groups (cf. B vs. C difference for balanced Unit #1 in Table 2).

Let us turn next to the *balanced* unit containing two negative and one positive valence (Unit #4, Table 1). According to the theory, if p *dislikes* o, q, and r, then p will tend to perceive these persons as *liking* each other. The relevant data for testing this hypothesis are

TABLE 2
BASIC DATA FOR BALANCED AND IMBALANCED UNITS

Unit[a]	Group	*N*	Number of Choices: Total possible	Chance expectancy (A)	Perceived by p (B)	Actual (C)	Significance of differences based on chi-square tests, 1 *df*: A vs. B	A vs. C	B vs. C
Balanced #1	I	22	132	18.9	75	59	.001	.001	.01
	II	16	96	19.2	51	37	.001	.001	.01
	III	15	90	19.4	54	44	.001	.001	.05
	IV	12	72	19.8	42	25	.001	.20	.001
	V	12	72	19.8	33	26	.001	.20	.10
Balanced #4	I	22	132	18.9	47	34	.001	.001	.01
Imbalanced #6	I	22	198	28.3	16	19	.02	.10	.50
Imbalanced #7	I	22	198	28.3	7	6	.001	.001	.70

[a] Cf. Table 1 for description of specific units.

compare the observed number of attributions. Consider, instead, chance expectancy, one of the baselines mentioned above. If perceivers behaved in a random manner,[2] the number of choices p would expect to exist among the three members he chose would be $(3^2 \times 2)/(N - 1)$.[3] If the contribution of each perceiver is summed, the expected value would be $N(3^2 \times 2)/(N - 1)$. These expected values apply to

[2] The assumptions inherent in the present chance model are these. First, independence is assumed between the different selections made by any single individual. Second, the selections of any *S* are taken to be independent from those of any other *S*. Third, an *S* may not select the same other person more than once. Finally, he may not choose himself.

[3] The expected value is computed as follows. In a group of *N* members, o, q, and r must each choose three persons out of $N - 1$ others. The probability that o will choose q is then $3/(N - 1)$. Hence, the probability that both q and r will be chosen by o is $2 \times 3/(N - 1)$. The same set of operations is applied to q and r, with the result that the overall number of choices expected by chance between the members of any triad is $3(2 \times 3/N - 1)$ or $2 \times 3^2/(N - 1)$.

[4] This fact is in line with Newcomb's (5) suggestion that systems should show "symmetry" whether defined objectively or viewed through the eyes of a specific perceiver.

available only for the 22-man group. The chance baseline here is identical to that described above for Unit #1. A comparison of the total number of attributions in the group with the expected chance value yields a chi-square significant well beyond the .001 level (cf. A vs. B difference for balanced Unit #2 in Table 2). Of greater interest, however, is the fact that the level of attribution again significantly exceeds actuality (cf. B vs. C difference for balanced Unit #2 in Table 2).

Finally, we turn to the two *imbalanced* situations (Units #6 and #7, Table 1), where data are available for the 22-man group only. According to Heider, if p likes o but dislikes q, then situations in which p perceives o as liking q should be extremely rare; correspondingly, if p dislikes o but likes q, p should very seldom perceive o as liking q. The analysis in the present case takes the form of counting the number of positive choices p perceives as directed by his three most liked to his three least liked companions, and the number of choices perceived as going in the opposite direction. Here the number expected by chance[5] for any perceiver is $3^3/(N-1)$, or $3^3N/(N-1)$ for the whole group (cf. Column A, Table 2, for imbalanced units). Chi squares for both imbalanced units (cf. A vs. B difference for imbalanced units in Table 2), show that imbalanced configurations occur significantly less often than would be expected by chance. The number of actual positive choices is also below chance expectancy, very significantly for #7 and at the .10 level for #6 (cf. A vs. C difference for imbalanced units in Table 2). The difference between perceived and actual values in these imbalanced cases yields negligible chi squares (cf. B vs. C difference for imbalanced units in Table 2). This last finding differs from that obtained for balanced situations, where, it will be recalled, attribution differed significantly from actuality. There is, however, a distinct "floor" effect in the imbalanced cases. While expected values are fairly low (28.3/198), actual values are even lower (19/198 and 6/198). Hence, there is very little "room" available for attributed values to fall significantly below the actual number of choices exchanged.

[5] The expected value is computed as follows. Consider a group of N members and randomly select any two triads whose members are o, q, r and o′, q′, r′. Each must choose three persons out of $N-1$ others in the group. The probability that o, for example, will choose q′ is $3/(N-1)$, and, hence, the probability of o choosing o′, q′, and r′ becomes $3 \times 3/(N-1)$. Given identical probabilities for q and r choosing o′, q′, and r′, the overall number of choices expected by chance directed from the three members chosen by p to the three members rejected by p is $3 \times 3 \times 3/(N-1)$ or $3^3/(N-1)$. This expected value is obviously also applicable to the other imbalanced case in which choices are directed from a rejected triad of members to members in a chosen triad.

DISCUSSION

The present work applies the Heiderian model to the perception of affective feelings among members of small well-acquainted groups. The results support Heider's hypotheses regarding the cognition of affective relations. This evidence differs from, and thus supplements, the empirical studies of Heider's theory undertaken by Horowitz, Lyons, and Perlmutter (3) and by Jordan (4). In the former study, the emphasis is upon the evaluations of the *acts* of liked and disliked others. Jordan's experiment, on the other hand, while supplying data on affective relations, deals with *S*s' pleasantness-unpleasantness ratings of *imaginary* interpersonal situations verbally described.

We considered four of the eight cognitive units generated by all possible combinations of Heider's L-relations. In the uniformly positive balanced unit, perceivers attribute more affective links than expected by chance among the three members they prefer most. More important, however, is the fact that the attributed number of choices is also significantly in excess of the actual number existing between chosen others. Hence, the cognitive disposition toward balanced units, while deviating from chance in the direction of veridicality, is sufficiently strong to be a source of some systematic perceptual error. This "overshooting of actuality" was also observed in the other balanced situation under investigation, where the perceiver assumes that those he dislikes tend to choose one another more than they actually do.

The two imbalanced units examined also conform to Heider's prediction. In both cases, these cognitive configurations have frequencies significantly below chance, pointing to the rarity of imbalanced cognitive structures. It should be noted, however, that this veritable absence of cognitive imbalance is in line with actuality. Apparently, this is a very difficult

state to maintain in the actual social microsystem as well as in the cognitive representation of this system. While the occurrence of imbalanced configurations is rare, the differential frequency of the two types of imbalance is worth noting. It appears to be cognitively more difficult to conceive that a person we like will choose one we do not like than it is to conceive that a person we do not like will choose one we like. The present evidence implies that a simple balance-imbalance dichotomy does not do complete justice to the data. Rather, as Cartwright and Harary (1) also point out, there are degrees of balance and imbalance. Given the possibility, with the present methodology, of distinguishing degrees of cognitive and actual balance and imbalance, individual differences in cognitive dispositions toward balance become accessible to research.

An adequate conceptualization of the problem of cognitive balance must deal with the degree of correspondence between cognitive systems and systems objectively defined, a matter relatively ignored in discussions of Heider's theory. Is it legitimate to speak of a strain toward cognitive balance or away from cognitive imbalance in those cases where the cognitive and objective systems are highly similar? Could we not explain the situation more simply in terms of the *S*s' *accurate* awareness of reality? The shortcoming of an accuracy explanation in the present context is that it fails to account for the fact that strain toward cognitive balance may manifest itself in the very choice (or rejection) behavior of a member. We have shown that the level of affective compatibility among members chosen or rejected by a person is already higher than that of a randomly assembled triad. This very fact implies that cognitive determinants are involved and supports the Heiderian principle of preference for balanced and aversion for imbalanced units.

Heider's theoretical system appears to be a most useful vehicle for exploring the problem of interpersonal feelings and perceptions. The present paper demonstrates that the hypotheses derived from the theory are confirmed when empirically tested in a small group situation and that new research problems emerge from the application of Heider's conceptions to this type of data.

SUMMARY

The perception of interpersonal preferences in small groups represents a problem area of immediate relevance to Heider's theory of cognitive configurations. In the present investigation, hypotheses derived from this theory are empirically tested and confirmed. Balanced cognitive units occur significantly in excess of chance and actuality baselines. Imbalanced cognitive units, on the other hand, appear significantly less often than would be expected by chance, though it should be noted that actual social units of this kind are also very rare. Theoretical and experimental implications of the findings are discussed.

REFERENCES

1. CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.
2. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
3. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.
4. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273–287.
5. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

Received April 22, 1957.

CASE REPORT

SOCIAL BEHAVIOR OF YOUNG CHILDREN IN THE *KIBBUTZ*[1]

HELEN FAIGIN

The Hebrew University, Jerusalem

THE system of collective education followed in the *kibbutzim* (collective settlements)[2] in Israel provides a unique opportunity for the study of the effects on personality development and social behavior of a pattern of socialization that stresses group living and the importance of peer relationships from the very beginnings of life. The study reported here is an exploratory investigation of the social relationships among children of toddler age growing up in this setting.

Within the system of collective education, the children are reared in groups from birth through adolescence. They live in their own houses apart from their parents and are cared for by a *metapelet* (children's caretaker) whose responsibility it is to care for and socialize the children. The role of the *metapelet* (plural, *metaplot*) changes with the age of the children and as the children grow older, teachers and the children themselves assume some of the roles which the *metapelet* had performed earlier. The children spend two to three hours every afternoon with their parents and most of the day on Saturday, the day of rest. Whereas the *metapelet* (especially in the early years) has responsibility and authority for the socialization of the children, the role of the parents is to be "good friends" with the children. The parents' role as disciplinary agents is negligible.

Toddlers' groups were chosen as the special focus of the study in order to determine whether under these special conditions of group-living social interaction between children and group identification can be learned in very young children. Most studies and observations of children in western societies indicate that children interact with one another to a very small degree and for short lengths of time until approximately the age of three years. There is little concept of group or group identification at a younger age. The children are oriented for the most part to themselves and to the significant adults in their environment. The development from a biological to a social being is a slow and gradual process. Do children living in groups from birth in this type of setting learn to interact with each other at an earlier age; do they begin to feel themselves as part of a group and develop close relationships with one another earlier than children reared in families? These were the primary questions investigated in this study. A secondary point of emphasis was the role of the *metapelet* and her relationship with the children.

METHOD

The Sample

The study was carried out in two settlements, in each of which three groups of toddlers were observed. The children ranged in age from 19 to 38 months at the time the study was begun. The groups were divided according to age, with an age span of five to six months within each group. All groups consisted of six children.

The two settlements had been in existence for approximately twenty years at the time of the study. The members of both *kibbutzim* are of European origin and have the same general educational level. Each of the *kibbutzim* belongs to a different *kibbutz* movement[3] and can be said to be typical to a certain degree of the educational philosophies and practices of that particular movement. The differences in educational orientation may be summarized in terms of the greater degree of deliberate educational stimulation and the more structured setting of *Kibbutz* S, as compared with the greater freedom allowed the child in *Kibbutz* M.

[1] This research was supported by the Committee for Research on Collective Education of Seminar Hakibbutzim.

[2] For description of the *kibbutz* setting and the system of Collective Education, see (2, 6, 7).

[3] Each *kibbutz* is a member of a *kibbutz* movement which coincides in general with a political party. Each movement has certain political, social, and educational ideologies which distinguishes it from the other movements. For fuller detail see Spiro (7).

Method of Observation

Observations were made on the children at regular intervals over a period of six months. Systematic observations were made on the aggressive and dependent behavior of each individual child[4] in various activities: free play, organized play, meal time, and getting up in the afternoon. In all, there were three hours of systematic observation made on each child in ten- or fifteen-minute time units. Specific time periods were allotted to each activity. Observations were made, also, upon each group as a whole in the above noted activities and, in addition, getting up in the morning, going to sleep in the afternoon and evening, on walks, at shower time, etc. In this way, the children were observed in all the daily routines. In addition, both formal and informal discussions were held with the *metaplot*, parents, and the person or persons responsible for organizing and directing educational practices with young children in each of the settlements.

Nature of the Data

Systematic observations of dependent and aggressive behavior of the children were made in terms of predetermined categories to be described, so as to permit statistical analysis of the frequency of different response categories.

Most of the observations were made by the writer. Help was obtained for a short period of time from two assistants in order to establish reliability of observation.[5] The reliability of observation of the various categories of aggressive responses ranged from 70 to 95 per cent, using the formula:

$$\text{per cent agreement} = \frac{\text{number of agreements of two observers}}{\text{number of agreements plus number of disagreements}}$$

Agreement on total aggression was 92 per cent. The reliability of the dependency categories ranged from 75 to 97 per cent. Agreement on total dependency was 90 per cent.

The data on group behavior were obtained from consistent rather than systematic observations. Full records were kept of the occurrences of responses of the various categories of social behavior observed. Special care was taken not to generalize from one instance of unusual or interesting behavior but to generalize only from types of behavior that occurred with some frequency over the entire period of observation. These data were used for descriptive purposes. All data on the behavior of the children are based on these two types of observation. Information was thus obtained on the day-to-day activities of the children, the patterns of interaction among them, and their interaction with the *metapelet*.

The data obtained from formal and informal discussions with *metaplot* and educators were used as background material and as a basis for assessing their attitudes toward the educational system and the problems, general and specific, that arise in taking care of very young children.

THE DAILY ROUTINE

The daily routine in the toddlers' house is quite similar in the two *kibbutzim*, but the daily program is somewhat more set and routinized in *Kibbutz* S than in *Kibbutz* M. The children wake up in the morning at about 6:00 or 6:30. There is a period of free play before breakfast usually inside the house. After breakfast there is again free play; the children usually play outside while the *metapelet* cleans up the house. After this the children usually go on walks around the settlement, visiting various work branches, go swimming in the summer, have their midmorning snack, play again around the house, and then have lunch. In *Kibbutz* S, almost every morning after the walk, or in place of it, the children spend about half an hour in concentrated activity known as "work": in general, some creative or constructive activity, such as

[4] There were six children in each group, but the individual child sample included only those children born in the *kibbutz*. This made a total of 31 children: five children from five of the groups, and six children in one of the groups.

[5] These assistants were seminary students, who were given special training in the observation and recording of aggression and dependency, including ten hours each of observing dependency and aggression in groups of children of similar age but not in the research sample, during free play activity. The data on reliability summarized here were obtained from ten hours of joint observation by the author and an assistant. Of the five hours that each assistant observed with the author, three hours of observation were made during the free play period. Observations were divided quite equally among the three groups of children.

drawing, building, or reading. Emphasis is placed on working well and quietly. This morning "work" period tends to become a routinized daily activity when the children are about 24–27 months old. In *Kibbutz* M, concentrated activity is an almost daily occurrence in the oldest (3-year) group and occurs once in a while in the younger groups. The concept of "work" is not employed and there is not as much emphasis on quiet concentrated activity.

The children eat lunch at about 11:00 A.M. and then go out and play in the yard while the *metapelet* cleans up. After this they shower and take their afternoon nap, which usually lasts around two hours. The children are washed, dressed, and given a snack, and then go off to their parents. In *Kibbutz* M, the children have their evening meal at this time. The reason for this will be explained below.

The children spend two to three hours with the parents and siblings. The parents (one or both) bring the children back to the children's house. In *Kibbutz* S, the children have their evening meal, and the *metapelet* puts them to bed. In *Kibbutz* M, the parents bring the children back to the children's house and put them to sleep. In both *kibbutzim*, the children sleep in the houses all night without any adult with them. A night watchwoman, who is assigned to the children's houses, walks about in the area and looks after any children that cry or have any particular problems.

THE ROLE OF THE *METAPELET*

The role of the *metapelet* of the toddlers includes a large number of duties and responsibilities; the *metapelet* functions as housekeeper, caretaker, socializing agent, and teacher. She does all the things that a mother usually does and has been delegated the responsibility and authority for rearing the children. Her evaluation as a *metapelet* is in terms of how well the children in her care develop. A close attachment exists between the children, and the *metapelet* and the children will often go to visit the *metapelet* who worked with them at this age when they are older and she no longer works with them.

The Metapelet as Housekeeper

One of the *metapelet's* duties is to keep the children's house clean and in order. She washes the floors, the dishes, makes the beds, takes care of the laundry and in general keeps the house in order. To this task the *metapelet* devotes more or less time depending upon the age of the children, upon her own interests, her efficiency as a housekeeper, and how she herself defines her role.

The Metapelet as Caretaker

Physical care of the children. The *metapelet* feeds, washes, bathes, and clothes the children. She takes care of them when they are ill, calls the doctor if necessary, does all that is necessary for their general health.

Emotional care of the children. This is the area around which many of the problems of collective education are centered. There is much discussion, both by parents and by *metaplot*, about whether or not the children get enough individual attention. Since the group is composed of six children, the problem of balancing individual needs against group needs often arises. There are special courses for *metaplot*, and many discussions are held about the emotional needs of the young child in an effort to help with the problem arising in this area. The *metapelet* usually tries to give the child as much attention as he demands; the children who demand more usually receive more. The *metapelet* often spontaneously hugs and kisses the children when she is bathing and dressing them, holds their hands when they go for a walk, and kisses them when she puts them to sleep. There are individual differences among the *metaplot* in "warmth" and "motherliness" which influence the amount of attention and affection that they express toward the children.

There exists some conflict about the *metapelet's* infringing on the parents' loving role. The *metapelet* often rationalizes her feelings of not giving the children enough individual attention by saying that they get enough from their parents. On the other hand, the parents tend to resent the child's strong attachment to the *metapelet*.

The Metapelet as Socializing Agent

Among the important duties of the *metapelet* is the socialization of the basic disciplines. Weaning is done by the mother, but after this the responsibility is largely the *metapelet's*. She teaches the children to eat by themselves,

toilet trains them, teaches them to wash and dress themselves, answers their questions about sex, permits and limits their expression of aggression, and in general makes demands upon them in all areas of behavior.

Eating. The children begin eating by themselves at about one year of age. At this time, little fuss is made about the child smearing the table or his face. There are very few feeding problems. Since the children tend to imitate one another, learning to eat in a group seems to make the process easy and quick.

Toilet training. Toilet training is begun when the child is about eighteen months old. The children sit down in a group together, each on his own pot. The company of the other children and the presence of the *metapelet* seem to make toileting one of the activities of the daily routine. In each of the youngest groups (under two years of age) there were two children who did not always make a bowel movement in the pot. All the children in the other groups were bowel trained. Most of the children, even in the youngest groups, usually urinated in the pot during the day; occasional accidents occurred when the child was involved in his play. Two children in the oldest groups (two-and-a-half to three years) occasionally wet at night, the others were all dry. In the middle groups (two to two-and-a-half years) most of the children occasionally wet at night, but none wet every night. In the youngest groups, (under two years) there were four children who wet every night, the rest, occasionally.

Sex training. There is no separation of the sexes in the groups. The children sleep, bathe, and go to the toilet together. The children masturbate sometimes when sitting on the pot, when standing about doing nothing, and occasionally when upset. The children masturbate each other occasionally. The *metapelet* usually pays no attention to the child's masturbation, but occasionally tries to distract him by getting him interested in some activity.

Independence training. The children begin to try to dress themselves by about the age of two years. There is no direct pressure for them to do so, but they are highly praised for their attempts in this direction. They eat alone at an early age, and in general are encouraged to do things by themselves. They are praised for their achievements. Emphasis is placed on "being big." There is somewhat more emphasis on independence in *Kibbutz* S than in *Kibbutz* M.

Control of aggression. The problem of aggression is discussed later in detail. In general, the *metapelet* does not interfere in quarrels between the children, only in cases when a child is getting badly hurt. The general attitude is that the children should have the opportunity for working through their quarrels by themselves.

Discipline. Physical punishment is taboo in the *kibbutz*. Discipline can be characterized, in general, in terms of explaining and trying to win the child over by suggestion or persuasion. A very effective technique is to separate the child from the other members of the group until he is ready to behave properly. Other members of the group often join in and suggest that the "erring" child be deprived of something or they tell him what to do.

The children are very cooperative in general with regard to the demands made upon them, and there are no real disciplinary problems in these groups. Disobedience and disciplinary problems arise much more frequently when the child is with his parents. The general atmosphere within the groups is permissive, and few demands for obedience are made upon the child. There is a tendency to conform to group demands and pressures.

The Metapelet as Educator

Education as used here refers to the active teaching and stimulating of the children as opposed to the subtle, barely conscious teaching of attitudes and relationships. The *metapelet's* role as educator is greatly stressed in *Kibbutz* S, whereas in *Kibbutz* M it is hardly considered a part of her role with these young children.

The *metapelet* in *Kibbutz* S also sees guiding and educating the parents in how to handle their children as an important part of her task.

THE CHILDREN'S GROUPS

The children's groups are described in terms of the general tendency toward group identification, the quality and types of interaction, and the effects of group living on the social development of the children.

Identification with the Group

There seems to exist a certain strong "we group" feeling in all of the children's groups. This was evident in their supporting each other whenever criticized or attacked by children from another group. All toys, furniture, and space within the children's house were referred to as "ours" as opposed to "theirs." There were many instances of out-group aggression, and the group united to support its own members. The children often did not permit children from other groups to play with their toys or to come into their house. Whenever a child was absent from a group activity, his absence was noted and mentioned. For example, when a group goes for a walk, even the youngest groups always see to it that all the children are present before going off, without any prompting from the *metapelet*. One of the favorite games of the children was playing on a carousel. They always made room for the members of their own group, but would not permit children from other groups to join them.

The concept of "we" and "ours" was well developed in the children by the age of two. All activities and property connected with the children's house were referred to by the children as "ours," whereas anything that a child had or did in his parents' room was referred to as "mine." A child might bring candy from his parents' room to the children's house and it was up to him whether or not he would distribute it; if he did, he usually did so with much ceremony. The general idea is that this is mine and I am sharing it with you. He does not have to share, however, and no pressure is put upon him to do so by the *metapelet*. The attitude toward things around the children's house, however, was quite different. Here, the expected thing was to share—everything is *ours*. Even when a child had a coveted toy which he did not want to give to another child—he would not say "this is mine"—he would say, "but I had it first."

The children in one group would often brag about the things they had done or the property they had in their house to children of another group. "We have a nice house and you don't," "we went for a walk and you didn't," and the children of the other group would retort in kind.

If a child was hit or teased by a child from another group, the other children in his group would always rise to his defense no matter what the reason was for the quarrel. The children in a given group would often adopt the same attitudes toward things: such and such food is good, such and such is not, such and such toy is good, such and such is not, such and such person is good, such and such is not.

When a child did not want to do something —for example, eat, help clean up, look at a book, in general to do the same activity as the rest of the group—the *metapelet* would usually say, "If you don't want to you don't have to, you can go out and play." This was reacted to by the children as punishment and they joined in the activity.

The children often imitated one another at mealtime and in many of the group activities. Often when the *metapelet* helped a child eat or dress, this child would tell her to help the other children also.

In *Kibbutz* S, this group orientation and identification was more verbalized both by the *metaplot* and by the children. The *metapelet* very consciously talked about the group, in many ways encouraged the development of the concept "we" and "ours." In *Kibbutz* M, this was done much less consciously, but the group feeling exists. A favorite game of the children in the oldest group in *Kibbutz* M, when the observer was present, was to ask her to write down their names. Each child would ask to have his name and the names of all the other children in the group written down. No child ever omitted the name of any of the other children in the group. In both settlements, the children always made sure that everyone was present at mealtime, and at all group activities, such as going for walks and concentrated activity such as reading, drawing, and making things out of plastalina. This making sure that everyone is present was more frequent and consistent in *Kibbutz* S.

The more conscious awareness and expression of group feeling in *Kibbutz* S than in *Kibbutz* M seems attributable mainly to two factors: (*a*) the greater general awareness and articulation in the *Kibbutz* S movement of inculcating the ideology of group living in the young children and (*b*) in *Kibbutz* M, the groups break up at the age of three to three-and-a-half to go to separate kindergartens and come

together again only in the second grade—whereas in *Kibbutz* S the group continues to stay together.

Social Participation in the Children's Groups

Much of the time the children spend playing separately, and yet together. Each child is involved in his own particular activity, but plays in the vicinity of the other children. Each child looked around from time to time to see if the others were present and would often point out to the other what he was doing. If any child was not present, the other children would inquire as to this child's whereabouts.

Even in the youngest groups the children often played games together such as peek-a-boo in a group or just two or three of them together. They often imitated one another, crawled on the floor, or pretended that they were dogs. Often two or three children would evolve some game which might go on as long as 15 minutes. In the youngest group, when they played in twos and threes, the specific children changed from activity to activity and from game to game. In the two older groups, certain friendships had formed and there were often two or three children who usually played together. There began to be a trend toward pairing off or forming smaller groups within the group. This was observed mostly during free play.

In the youngest groups, "play together"—aside from general group activities such as going for walks or riding on the carousel—usually involved looking at things such as grass or insects together—and laughing, playing peek-a-boo, imitating animals, and occasionally building things together. In the middle groups, the children often build things together and play more involved games such as picking and exchanging flowers, exchanging toys, and building canals together out of sand and water. In the oldest group, the children built things together, and there was also quite a bit of imaginative play in which the children took on roles of adults. The most popular of these games in the oldest groups in both *kibbutzim* was playing mother, father, and baby. The same game would be continued by the same group of children day after day, during some of the free play time. Usually, the same child would play the same role each day; sometimes there would be shifts in the child who played a particular role. It is interesting to note, however, that this imaginative play centered on the family, not on the children's group and the *metapelet*. Occasionally, the children played doctor and similar games, but the family group was most popular. Further investigation of this problem in terms of analysis of the children's play might prove rewarding in getting at the kinds of relationships these children have with their families as compared to children growing up in private families.

Group Values

One of the important values expressed by the children both in their behavior and verbally in all the groups in both settlements is sharing and taking turns. This was apparent especially with regard to food, and playthings such as paper, crayons, and plastelina. The children knew that they would each get some and that they had to wait their turn. How much the children learned in this area was especially evident when a new child came into the group who had not learned to share and wait his turn. The other children explained to him and often pounced on him when he did not behave properly.

In *Kibbutz* S there was also much stress put upon being big and strong and independent, and on doing "constructive" things—"working" quietly and making things. The "big" and independent child can take care of himself, can dress himself, and doesn't cry when he gets hurt. In *Kibbutz* M, children were praised for these things, but there was little stress placed on them.

Values expressed with regard to fighting and displaying affection are discussed later.

The children's groups, even the youngest, play an important part in setting standards for behavior and in controlling the behavior of group members. Often the children's statements of encouragement or discipline were direct imitations of the words of the *metapelet*; sometimes they were inventions of the children themselves. Whichever they were, they were used at the appropriate times and carried the weight of sanctions. The children to whom the statements were addressed responded in accordance with them.

In the youngest groups in both settlements, most statements of social control referred to eating and toilet training. Some times they

also referred to matters of disobedience and aggression. For example, when one of the children made a bowel movement on the floor the other children in the group looked at him and they all said, "Haggai did a BM, Haggai did a BM, not on the floor, Haggai, not on the floor." At another time the children were sitting around the table eating soup. One of the children said the soup was hot, so another child said to her "It's not hot, taste it and you'll see, it's not hot." At times, this control took the form of action rather than words. One of the children had been hitting another child; the second child was crying. All the rest of the children in the group came over and hit the first child.

In the older groups in both ***kibbutzim*** this control covers all areas of behavior. Sometimes the older children told the younger children how to behave. Below are several examples of statements used by the children in the two older groups to control each other's behavior. First, statements referring to eating:

One of the children isn't eating. Another child says, "Why aren't you eating—I'll feed you." He held him by the arm and fed him.

When one of the children did not want to eat her soup, another child said to her, "It will cool off soon, then you should eat it, when it cools off."

One of the children wanted the observer to write his name, but he hadn't yet eaten his lunch. One of the other children said, "She won't write it for you until you eat." All the other children joined in and told him that first he had to eat and then she (the observer) would write his name.

One of the children wanted to play with nails. Two of the other children told the *metapelet*, "Don't give him any nails yet, he hasn't finished eating."

Statements referring to aggression:

Two children were standing in the house looking out the window. They saw another child hitting a baby. One said to the other, "One mustn't hit babies." Then he called out to the child outside, "You mustn't hit the baby."

One child said to another, "I'm going to hit you." The second child said, "Why do you want to hit me?" The first child then said, "I won't hit you" and hit the table instead.

One child hit another. A third child came over and hugged the child who had been hit and asked, "Who hit you?" He then told that child that he must not hit the other child again.

One child hit another. The second child said, "Why are you hitting me? I'm not hitting you." The first child stopped hitting.

One child hit another. A third child told the child who had hit, "How many times do I have to tell you not to hit other children."

Social Control

Statements referring to being careful:

One of the children put a pin in his mouth. Another child told him, "You mustn't put a pin in your mouth—you'll get a sore." She took the pin away from him.

On sharing:

Two children were fighting over who was to sit by the steering wheel in an auto. A third child said: "Giladi, that belongs to all of you, to Yisrael too, he can sit there, too, both of you can." They both sat down beside the steering wheel.

The children from a very early age have thus taken over the values of the culture in which they are being reared and use these actively in setting standards for each other's behavior. Not only are these social control statements used but they carry weight in controlling behavior of the children. So the group operates as a socializing agent upon its members from a very young age. In many instances the forms of social control expressed and exercised by these young children seemed to have the quality of "punishment by reciprocity" and to carry the notions of "equality" and "solidarity" as these concepts are used by Piaget (4). These children are much younger than the age at which the idea of distributive or equalitarian justice are thought to develop, but perhaps the conditions of group living tend to induce the development of cooperation, equality, and group solidarity at a very early age. A more focussed study of the methods of social control used by very young children would seem highly desirable.

It is interesting to compare these observations with the use of rules by three-year-old middle-class nursery school children studied by the same investigator (1) in a small city in the United States. Two interesting differences may be noted. The American three-year-olds tended to use more rules about property, that is, about toys and play space, than rules of interaction, how to behave. In the ***kibbutz*** group, the children although younger tend to use more rules about social interaction than property. This difference would seem to follow from the fact that the ***kibbutz*** children live and interact with one another to a much greater extent than do nursery school children. The other interesting difference concerns the nature of the property rules. The American

nursery school children used rules referring to this being "mine"—"I had it first, you can't have it"—whereas the *kibbutz* children's property rules refer to sharing, "This is ours, this belongs to all of us." The social control methods used by these very young children thus appear to reflect the values of the society in which they grow up: one a society in which private property is highly valued, the other in which there is communal ownership and no private ownership.

Relations Between the Children

The affective ties between the children are very strong. They do not approach the intensity of relations observed among siblings outside the *kibbutz*, and yet they are much more intense than that of friendships among other Israeli children of this age who play together. In spite of the fact that much of the time the children play by themselves, they almost always play in the vicinity of the other children and are carefully attuned to the movements of the other children. They are aware of the presence and absence of the other children, know what they are doing, respond immediately to the crying and fighting of the other children.

Many of the *metaplot* and parents have some doubts about whether or not children of this age should be living in groups. They talk about the "struggle for existence" which goes on among the children. They say that there is too much fighting and that the strong, more aggressive children hold the power and can get what they want, whereas the weaker, less aggressive children have a hard time holding their own and have to give in to the stronger. This is only partially the case. There is much fighting and crying and the stronger children often do hit the weaker children and take the toys away from them. On the other hand, the children often comfort each other when they are upset or crying. When a child is crying because he has been hurt or because a toy was taken away from him, very often the aggressor will come over and pat the child, tell him he will not hurt him again and perhaps return the toy he has taken or give him another in exchange. If the child who has been the aggressor does not comfort the hurt child, other children in the group will do so.

In each of the two youngest groups there was one extremely aggressive child. These children very often hit the other children and took toys from them with seemingly no provocation, at least no provocation apparent in the situation. The other children tended to avoid them, and from time to time would gang up on the aggressive child. In the two older groups in each of the *kibbutzim* there was no child who could be called extremely aggressive, in the sense of showing behavior similar to the two children in the youngest groups.

In none of the groups was there a domineering or bossy child who would try to tell all the others what to do. During the course of the observation, a new child came into one of the groups who tried to boss the other children around. She quickly learned that she could not do this. The other children either paid no attention to her or fell upon her as a group and beat her. In a very short time this behavior disappeared.

There was no child in any of the groups who could be called the leader. At different times different children played leadership roles in given activities, but no child played this role consistently. Leadership was always in the hands of the *metapelet*. Often when older children would come to visit, an older sibling of one of the children in the group, this older child would take a leader role, organize some activity and the children in the group would follow this leadership.

SYSTEMATIC OBSERVATIONS ON AGGRESSION AND DEPENDENCY

Aggression

It is difficult to give a general picture of the frequency or types of aggression expressed by the children in the two settlements. Systematic observations were made of the frequency of various types of aggression expressed by the children during free play. Each child was observed at four different periods of 15 minutes each. Nothing can be said about whether children in *kibbutzim* display more or less aggression than children growing up in private families, since there are no comparative data. The data have been analyzed to compare the children in the two *kibbutzim*, with special emphasis on the different educational practices, and to compare the different age groups within each *kibbutz*.

TABLE 1
MEAN AGGRESSIVE RESPONSES FOR THREE GROUPS OF CHILDREN IN EACH *Kibbutz*

	Mean Number of Responses per Child During Four 15-Minute Periods of Observation							
	Kibbutz M				*Kibbutz* S			
Category	Grp. A[a] ($N = 5$)	Grp. B ($N = 5$)	Grp. C ($N = 5$)	All Grps. ($N = 15$)	Grp. A ($N = 6$)	Grp. B ($N = 5$)	Grp. C ($N = 5$)	All Grps. ($N = 16$)
Physical aggression	4.2	3.4	1.2	2.9	2.5	2.4	2.6	2.5
Instrumental aggression	1.0	3.4	1.8	2.1	4.3	1.0	1.0	2.2
Defense of toys	0.2	0.8	1.4	0.8	0.5	0.8	3.0	1.4
Defense of self	0.2	0.8	0.6	0.5	0.2	1.2	1.2	0.8
Aggression to objects	1.2	2.0	0.0	1.0	2.8	2.0	3.8	2.5
Verbal aggression	0.0	0.0	2.0	0.7	0.0	0.2	0.8	0.3
Aggression to smaller children	0.0	0.8	2.2	1.0	0.0	0.4	2.8	1.0
Disobedience	0.4	1.0	0.4	0.6	0.3	0.4	1.8	0.8
Total aggression	7.2	12.2	9.6	9.7	10.7	8.4	17.0	11.9
(*SD*)	(6.2)	(6.8)	(7.2)	(6.6)	(11.2)	(5.8)	(11.5)	(10.6)

[a] In each case, Group A is composed of the youngest children, Group C of the oldest.

The aggression shown by the children was broken down into various types of aggression. The data on the frequency of all types of aggression recorded are shown in Table 1.

Physical aggression. Physical aggression is defined in terms of hitting for the purpose of hitting or hurting. There were no statistically significant differences between the children in the two ***kibbutzim*** in the frequency of use of physical aggression as can be seen from Table 1. Within the ***kibbutzim*** the only significant difference between age groups was found in *Kibbutz* M in which the children in the oldest group exhibited significantly less physical aggression than the children in the middle age group. The highest frequency of expression of physical aggression was by a child in the youngest group in *Kibbutz* M, one of the very aggressive children previously described, for whom 15 units of physical aggression were recorded in the hour's observation. For many of the children, 6 out of 31, no physical aggression was recorded during the hour period, and for nine children only one instance of physical aggression. Thus, about half the children (15 out of 31) expressed very little or no physical aggression.

Instrumental aggression. Instrumental aggression is defined in terms of hitting or pushing in order to attain some other goal, such as a toy, play space, or attention. Table 1 shows the frequency of expression of instrumental aggression. There were no significant differences between the ***kibbutzim*** or between the various age groups. The highest frequency of instrumental aggression was recorded for one of the previously described aggressive children, in the youngest group in *Kibbutz* S, 20 units. For 13 children, no instances of instrumental aggression were recorded, and for five children only one instance of instrumental aggression was recorded. Thus, for 18 out of 31 children there was practically no instrumental aggression expressed.

Other forms of aggression. The frequency of occurrence of the other categories of aggression (defense of toys, defense of self, aggression toward objects, verbal aggression, aggression toward younger children, and disobedience) is so small that not much can be said about it except to note its lack of occurrence.

Total aggression. Total aggression includes all forms of aggression, the total scores for the various types of aggressive responses. There were no significant differences in the frequency scores of total aggression (Table 1) between the two ***kibbutzim***, between boys and girls, nor among the different groups. There were large individual differences in the frequency of occurrence of aggressive responses as reflected in the large *SD*'s, presented in Table 1, with a range from 1 to 39 in the frequency of total aggression.

These wide individual variations would seem to indicate that the instigation to aggression is due to factors other than or in addition to the frustrations inherent in group living. It was observed that not only did the individual

metaplot respond differently in general to the aggressive responses of the children and by their patterns of relationship instigate more or less aggression, but each *metapelet* responded differently to each individual child when he responded aggresively. Some children were encouraged to respond aggressively, some discouraged from this, and some "just got away with it." Although the *metaplot* tend to practice a "hands off" policy, they were frequently inconsistent in their reactions to the aggressive responses of individual children. The presence or absence of the *metapelet* at the time of occurrence of aggressive responses adds to the factor of inconsistency. In addition, the children themselves respond differentially to one another in reaction to aggression directed toward them. The family constellation, too, probably has an important effect on the child's tendency to respond aggressively. It was noted, however, that children who are less aggressive in the group are more aggressive in the parents' room and vice versa.

Dependency

The dependency behavior expressed by the children was observed and recorded systematically as was aggressive behavior. Under the general heading of dependency the following types of behavior were recorded: crying, showing affection toward adults (other than the *metapelet* and observer), showing affection toward the observer, showing affection to the *metapelet*, showing affection toward children, instrumental dependency, i.e., asking help to do something, seeking attention or approval for making or doing something, and thumbsucking. The frequencies with which these responses occurred in the various groups in the two *kibbutzim* may be seen in Table 2.

Expressing affection was categorized under dependency since it is also a means of seeking attention and affection. Affection is expressed by responses such as touching, holding, or being near. The nature of these responses inevitably induced a response of a similar type from the other person—thus the child's giving and seeking attention could not be separated.

Dependency toward adults. The one consistent difference that appeared in all the categories of dependent behavior (except affection toward children and displaying achievements) was that between the children of the youngest groups in the two *kibbutzim*. In total dependency toward adults, the children in the youngest group in *Kibbutz* S showed higher scores than the children in the youngest group in *Kibbutz* M.[6] This same type of difference was found in the separate categories of affection toward adults, toward *metapelet*, toward observer (.10 level of confidence) and instrumental dependency (.10 level of confidence). In contrast, the children in the youngest group in *Kibbutz* M cried more frequently (.10 level of confidence) and sucked their thumbs more frequently (.10 level of confidence) than did the children in the youngest group in *Kibbutz* S.

These apparently inconsistent findings probably can best be understood in terms of the differences in educational orientation in the two *kibbutzim*. It will be remembered that *Kibbutz* S was described as stressing educational stimulation more at an early age, whereas *Kibbutz* M tends to stress a great deal of freedom and permissiveness, letting the child develop at his own rate. One of the consequences of this difference is that the *metapelet* in *Kibbutz* S spends more time with the children during free play and hence they direct their dependency responses toward her. In contrast, the children in *Kibbutz* M tend to cry more unheeded and suck their thumbs since the *metapelet* is not around. There is less opportunity to direct dependency responses toward adults. In general, the children in the youngest group in *Kibbutz* S seemed to develop at a faster rate than the children in *Kibbutz* M. They seemed to be more aware of adults, more aware of their surroundings, played more, and spoke earlier than the children in *Kibbutz* M. These differences in rate of development seem to level off in the two older groups where the same differences do not occur.

Differences were found between the oldest groups in the two *kibbutzim* in two categories: displaying achievements and instrumental dependency. Here the children of *Kibbutz* M more frequently displayed this behavior than the children of *Kibbutz* S, contrary to what might have been expected. A possible explanation is that the *metapelet* of the oldest group

[6] All differences mentioned are statistically significant at the .05 level of confidence using the *t* test, unless otherwise stated.

TABLE 2
MEAN DEPENDENCY RESPONSES FOR THREE GROUPS OF CHILDREN FROM EACH *Kibbutz*

Category	Mean Numbers of Responses per Child During Four 15-Minute Periods of Observation							
	Kibbutz M				*Kibbutz* S			
	Grp. A[a] ($N = 5$)	Grp. B ($N = 5$)	Grp. C ($N = 5$)	All Grps. ($N = 15$)	Grp. A ($N = 6$)	Grp. B ($N = 5$)	Grp. C ($N = 5$)	All Grps. ($N = 16$)
Crying	10.6	3.4	2.8	5.6	4.7	4.0	2.4	3.8
Affection to adults	1.8	7.2	6.8	5.3	5.8	9.0	8.0	7.5
Affection to observer	8.0	9.0	14.2	10.4	21.3	8.8	6.4	12.8
Affection to *metapelet*	0.8	4.2	6.4	3.8	2.8	9.4	8.4	6.6
Instrumental dependency	0.4	0.6	5.8	2.3	1.3	1.6	1.2	1.3
Displaying achievements	0.0	0.6	2.8	1.1	0.0	2.6	0.6	1.0
Total dependency to adults	21.6	25.0	39.0	28.5	35.8	35.4	27.0	32.9
Affection to children	1.0	5.6	4.0	3.5	1.3	2.0	4.0	2.4
Thumbsucking[b]	7.2	6.2	2.0	5.1	1.5	6.0	1.2	2.8
Total dependency	22.6	30.6	43.0	32.1	37.3	37.5	31.0	35.3
(*SD*)	(10.4)	(5.0)	(17.4)	(15.2)	(9.5)	(20.6)	(19.0)	(16.9)

[a] In each case, Group A is composed of the youngest children, Group C of the oldest.
[b] Not included in total dependency score.

in *Kibbutz* S usually praised the achievements of the children without their asking for it. The difference between the two groups on instrumental dependency is more easy to understand. As noted previously, the *metapelet* in *Kibbutz* S encourages and demands more independence on the part of the children than the *metapelet* of *Kibbutz* M. She urges them to do things by themselves rather than to ask for help. She often refuses help when the children ask for it, telling them instead to try to do whatever they wanted by themselves.

Dependency toward children. There is a general tendency for the amount of affection and dependency expressed toward children to increase with age. That is, the children in the youngest groups showed affection to other children less frequently than the children in the older groups. There were differences between the two *kibbutzim* only in the middle groups. The children of the middle group in *Kibbutz* M showed more affection and dependency toward other children than the similar age children in *Kibbutz* S. It is not clear why this difference should appear in the middle groups and not in the oldest groups. When the amount of affection expressed toward children was compared with the amount of affection expressed toward the *metapelet*, there were no differences found except in the middle group in *Kibbutz* S. Here the children expressed affection toward the *metapelet* significantly more frequently than toward other children. We would expect to find differences in the youngest groups, but this is not the case.

Relationship Between Dependency and Aggression

In terms of overall scores, the children in both *kibbutzim* expressed significantly more dependent than aggressive responses ($t = 6.95$). There is a significant negative relationship ($r = -.73$) between the frequency scores of aggression and dependency. That is, the children who are more aggressive are less dependent, and the children who are more dependent are less aggressive. Most studies of nursery school children indicate a general low but positive relationship between dependency and aggression. Lois Murphy (3) in a classical study on sympathy and aggression in young children finds a positive relationship between these two types of responses. In a study reported by Sears *et al.* (5) there tends to be a low positive correlation between these two types of responses. These studies were made on older children than those observed here. Perhaps at this age dependency and aggression are alternative responses to frustration and the children who have been successful in using one or the other type of response tend to persist in its use. If we look at each group separately we note that in all groups, except the middle group in *Kibbutz* M, there is a

negative correlation between total aggression scores and total dependency scores. The correlations range from −.09 to −.72. In the middle group in *Kibbutz* M, there is an unexplained positive correlation (+.82) between the frequency of occurrence of these two types of responses.

When analysis is limited to the aggressive and dependent responses expressed toward other children, there is no relationship (r = +.12) between the occurrence of each of these responses. In the groups taken separately, however, there is, in four out of the six cases, a high positive correlation (.55–.93) between the expression of aggression and dependency toward children. The children show aggression toward each other with significantly greater frequency than they show dependency (t = 2.00), in contrast to the findings with respect to total scores. Not unexpectedly, the children thus tend to respond to other children more frequently with aggression and toward adults with dependency. In the groups considered separately, the children in the youngest group in each *kibbutz* are more frequently aggressive toward one another than they are affectionate, whereas in the two older groups the frequency of these two responses does not differ significantly . At the younger ages, the children are probably more affectionate toward adults as the source of their main satisfactions and care, and are aggressive toward other children since other children do not satisfy their needs; if anything, their peers are probably more likely to frustrate and interfere with need satisfaction. The process of learning to be affectionate toward other children is presumably a more complicated process involving the development of identification with one's peers and of capacities and skills of helping one another out.

SUMMARY AND CONCLUSIONS

An exploratory study of the social development of children of toddler age growing up in *kibbutzim* has been described. Since these children live together in groups practically from birth, the setting is propitious for studying the extent and limits of social interaction among very young children. A secondary point of emphasis was the role of the *metapelet* (children's caretaker) and her relationship with the children.

The sample consisted of three groups of children from each of two *kibbutzim*. The children ranged in age from 19 to 38 months at the time the study was begun. Observations were made on the children for a period of six months. Systematic observations were made on the individual children, general observations were made on each group, and discussions were held with the *metaplot* and educators.

The first part of the report deals with a description of the daily routine of the children. Differences between the *kibbutzim* in degree of deliberate educational stimulation and of freedom vs. structure in the setting were seen as having consequences in the relative amount of time the *metaplot* from the two *kibbutzim* spend with the children.

Accompanying differences in the behavior of the children in the two *kibbutzim* were noted mainly in the youngest groups. The children in the youngest group in *Kibbutz* S showed more dependency responses toward adults, whereas the children in *Kibbutz* M tended to suck their thumbs more often and to cry more frequently. The youngest children in *Kibbutz* S seemed to develop at a faster rate in terms of being more aware of their surroundings, more aware of adults, playing more, and speaking earlier. These differences tended to level off in the older groups where no differences of this sort were noted. The children in *Kibbutz* S tended to verbalize group identification and being "big" more than the children in *Kibbutz* M, but no differences were noted in other behavioral indices of these factors except that in *Kibbutz* M the children in the oldest group had higher instrumental dependency scores than the children in *Kibbutz* S.

The role of the *metapelet* is discussed in some detail, in terms of her role as housekeeper, caretaker, and socializer of the children.

The main body of the study is concentrated on the relations among the children. There is a strong group identification, frequent use of the terms "we" and "ours" as against "theirs." Competition exists between groups rather than within groups. Each group defends its members from attack by other groups. There is not a very high degree of social participation among the children, especially in the youngest groups, but the children always play in the presence of the others, and the absence of any child is noted by the other children. The children tend to control each other's behavior

by telling each other how to behave. There is a tendency to conformity to group standards. Problems of discipline are negligible. Treatment of sex play and masturbation is very permissive. Toilet training and eating habits seem to be learned with a minimum of conflict.

There are wide individual differences in the frequency of aggressive responses. The frequency of dependent responses is significantly higher than the frequency of aggressive responses, and children who are aggressive tend not to be dependent and vice versa. Children tend to be more aggressive than affectionate toward each other, especially in the youngest groups, and to be as frequently affectionate as aggressive in the older groups. There is a positive relationship between aggression and affection expressed toward other children.

The results of this study indicate, then, that group identification, sharing, and group control of individual behavior can be learned by very young children under conditions of group living. Caution is in order in generalizing from these results in view of the lack of comparative data on similar age children growing up in private families, and of the fact that children in only two *kibbutzim* were studied.

REFERENCES

1. FAIGIN, HELEN. The frequency and types of rules used by preschool children in their social interaction. Unpublished master's thesis, State Univer. of Iowa, 1948.
2. IRVINE, ELIZABETH E. Observations on the aims and methods of child rearing in communal settlements in Israel. *Hum. Relat.*, 1952, **5**, 247–275.
3. MURPHY, LOIS B., *Social behavior and child personality*. New York: Columbia Univer. Press, 1937.
4. PIAGET, J. *The moral judgement of the child*. New York: Harcourt, Brace, 1932.
5. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.
6. SPIRO, M. E. Education in a communal village in Israel. *Amer. J. of Orthopsychiat.*, 1955, **25**, 283–293.
7. SPIRO, M. E. *Kibbutz: Venture in Utopia*, Cambridge, Mass: Harvard Univer. Press, 1956.

Received April 27, 1957.

CRITIQUE AND NOTES

TWO DETERMINANTS OF CONFORMITY: AGE OF INDEPENDENCE TRAINING AND n ACHIEVEMENT[1]

ALLEN M. KREBS

University of Michigan

McCLELLAND (2) has derived a measure of achievement goal learning, n Achievement, and has demonstrated it to be related to conformity. Winterbottom (4) constructed a childhood training schedule to measure the age of self-reliance training; she reported it to be related to a measure similar to McClelland's n Achievement. There has been no attempt made to determine the relationship of Winterbottom's measure to conformity behavior nor the extent to which *both* Winterbottom's scale and McClelland's n Achievement are capable of discriminating between high and low conformers. The formulations of Riesman (3) suggest that both factors in relatively strong operation would render individuals less prone to change their opinions because of external social pressure.

Three hypotheses were derived in an attempt to determine the importance of the two measures:

I. The more intense the orientation toward generalized achievement goals (n Achievement) the more resistant individuals are to opinion change.

II. The earlier the occurrence of independence training during childhood, the more resistant individuals are to opinion change.

III. n Achievement and independence training operate in a synergistic fashion i.e., individuals with low n Achievement and late independence training conform to a greater extent than under any other combination of the two factors.

METHOD

Thirty-six subjects (*S*s), undergraduate males at the University of North Carolina, were employed in this experiment. n Achievement data were collected in the prescribed manner (2); the measure of age of independence training was derived by modifying Winterbottom's questionnaire (originally designed for use with children) to serve as a self report for *S*s of college age.

The following examples of Winterbottom questionnaire items discriminated well between early and late independence training:

——— ——— To show pride in your ability to do things well.

——— ——— To try hard things for yourself without asking for help.

The first space was to be checked if *S* felt that the item was pertinent to his parents' demands; in the second space *S* placed an estimate of the age at which the behavior was demanded of him.

In the situation designed to measure conformity, two *S*s were instructed to remember as many objects as possible of those that were to appear on five successive presentations of a set of two slides. The objects on one slide differed from those on the other—one slide purported to represent objects found inside one town, the other to represent objects found within another town in a context of military intelligence. Since there were a great number of objects on the slides and since the duration of exposure was short, considerable ambiguity existed concerning the precise nature of the objects presented in either. *S*s' scores for conformity were based on the number of objects each *S* was persuaded to indicate had occurred on the "Town A" slide when in fact they had been presented on the "Town B" slide, or vice versa.

Each *S* checked a list containing the names of various objects and indicated in a space next to each object whether it was in Place A or Place B. The *E* then substituted bogus lists for the genuine ones and returned these to the *S*s. Each bogus list purported to represent the responses of the other *S*, and disconfirmed four items of his own list. Conformity scores could range from 0 to 4, depending upon the extent to which *S* accepted the bogus report.

TABLE 1

SUMMARY OF CONFORMITY, n ACHIEVEMENT, AND INDEPENDENCE TRAINING SCORES FOR THE FOUR GROUPS OF SUBJECTS

Group	Conformity		n Achievement		Independence Training	
	M	*SD*	*M*	*SD*	*M*	*SD*
Early Independence training High n Achievement	.600	1.075	8.7	2.28	6.92	.169
Early Independence training Low n Achievement	.750	.887	2.2	2.11	6.87	.433
Late Independence training High n Achievement	.875	.991	8.9	4.65	8.62	.679
Late Independence training Low n Achievement	2.40	1.350	3.3	1.90	10.01	2.236

[1] This study was conducted in part by the Organization Research Group of the Institute for Research in Social Science, University of North Carolina. The activities of the Organization Research Group are supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855(04)]. The research is based upon the author's Master's thesis in the Department of Psychology at the University of North Carolina conducted under the direction of Dr. June Chance and with the gratefully appreciated assistance of Drs. deCharms, Thibaut, and Mellinger.

Depending upon whether the *Ss*' scores for age of independence training and for n Achievement fell above or below the median values of the two measures, they were assigned to one of four cells: early or late independence training in combination with high or low n Achievement. Presented in Table 1 are summary descriptive statistics of the Winterbottom, n Achievement, and conformity measures. An analysis of variance was employed to determine the significance of differences in conformity among cells.

A q test, discussed by Dixon and Massey (1, p. 152, 169), was carried out to determine whether the low n Achievement, late independence training cell indicated greater conformity than the other cells combined proportionately, i.e., weighted one third each.

RESULTS

Results of the analysis of variance of the conformity scores are presented in Table 2. A biserial correlation of .214 was found between measures of n Achievement and independence training. This relationship is not statistically significant ($p > .10$), in contrast to expectations derived from McClelland's formulations.

The significant relationship demonstrated between n Achievement and conformity serves as replication of McClelland's earlier findings. The relationship indicated between independence training and conformity substantiates the second hypothesis.

Hypothesis III was supported. The q test indicated that the greatest conformity occurred under conditions of low n Achievement, late independence training ($p < .05$). n Achievement and relative age of independence training in concurrent operation seem thus to have a synergistic effect regarding conformity behavior. When an individual falls in the lower half of both measures, he tends to conform to a greater extent than when in the lower half of either measure alone.

TABLE 2

ANALYSIS OF VARIANCE OF CONFORMITY SCORES CLASSIFIED BY THE VARIABLES n ACHIEVEMENT AND AGE OF INDEPENDENCE TRAINING

Source	Sum of Squares	*df*	*F*	*p*
Independence training	.9264	1	6.728	$<.05$
n Achievement	.7013	1	5.093	$<.05$
Interaction	.4727	1	3.433	N.S.
Error	4.4072	32		

REFERENCES

1. DIXON, W. J., & MASSEY, F. J. JR. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.
2. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. RIESMAN, D., with GLAZER, N. & DENNEY, R. *The lonely crowd.* New York: Doubleday, 1955.
4. WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Univer. of Michigan microfilms, 1953, Publ. No. 5113. (Abstract)

Received February 15, 1957.

THE SCALE VALUES OF STATEMENTS OF ATTITUDE TOWARD THE LAW AFTER A 25-YEAR INTERVAL

JOHN E. MILHOLLAND

University of Michigan

THE VALIDITY of any measurement depends in part on the constancy of the measuring instrument. In the domain of attitudes, measurement of change or of comparative standing of groups, requires that the scale mean the same when used at different times and/or with different groups. No direct test of this requirement is available, of course, but an indirect approach is illustrated by the data reported here.

In May of 1956, 37 students in a graduate class in psychological statistics were assigned an exercise in scaling by the Method of Successive Intervals (1). The attitude statements were taken from the Scale of Attitude Toward the Law (4) constructed by Daniel Katz at Chicago in 1930. As the result of an afterthought, the scale values of the items on the two occasions were compared. The product-moment coefficients of correlation were .98 for Form A of Katz's scale, .99 for Form B. There are 20 statements in each form of the scale.

These high correlations are particularly interesting because the 1956 procedures did not duplicate those of 1930. Katz used the Method of Equal-Appearing Intervals (5), rather than Successive Intervals; and 11 piles, whereas the 1956 scaling was done with 9 piles. The 1956 data were re-analyzed[1] by the Method of Equal-Appearing Intervals, and the correlations between item scale values were again .98 for Form A, .99 for Form B. The correlation between the two solutions for the 1956 data was .99 for each form.

[1] Dr. Sam Komorita suggested the re-analysis and carried it out.

Apparently, the 1956 perception of a University of Michigan graduate student as to favorableness of a statement toward law enforcement is not different, in a statistical sense, from that of a 1930 University of Chicago graduate student. Also, the close correspondence between solutions by the Method of Successive Intervals and the Method of Equal-Appearing Intervals makes it seem hardly justifiable to use the more elaborate procedures of the former. These generalizations, however, might not apply to other kinds of attitudes—it may be that the scale of Attitude Toward the Law is peculiarly stable. Hovland and Sherif (3), however, report a correlation of .96 between values of Attitude Toward the Negro (2) items scaled about 20 years apart.

REFERENCES

1. EDWARDS, A. L. Psychological scaling by means of successive intervals. *Psychometric laboratory report.* Univer. of Chicago, 1951, No. 69.
2. HINCKLEY, E. D. Attitude toward the Negro. In L. L. Thurstone (Ed.), *The measurement of social attitudes.* Univer. of Chicago, 1930.
3. HOVLAND, C. I., & SHERIF, MUZAFER. Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.
4. KATZ, DANIEL. Attitude toward the law. In L. L. Thurstone (Ed.), *The measurement of social attitudes.* Univer. of Chicago, 1931.
5. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Univer. of Chicago, 1929.

Received December 17, 1956.

MOTIVATIONAL FACTORS IN THE VIOLATION OF A PROHIBITION

DON L. KIMBRELL AND ROBERT R. BLAKE

University of Texas

CONFORMING to a sign prohibition has been shown to be greater when the sign is "strong" than when it is "weak" in a study by Freed, Chandler, Blake, and Mouton (1). Secondly, more people conform when others conform to a restriction and fewer with knowledge of violations by others (1). Grosser, Polansky, and Lippitt have similar results or violation of a command (2). The power to change the frequency of compliance has, in addition, been shown to be determined partly by "who the others are," with Lefkowitz, Blake, and Mouton (3) investigating this factor in traffic light violations. The present paper extends previous work by relating strength of motivation to act against a prohibition to the social properties of the prohibition situation.

EXPERIMENTAL DESIGN

The experiment was presented to 90 college *Ss* as a study of taste preference.

Description of the Test Situation

Two *Ss*, an accomplice and an informed *S*, came to the laboratory at the same time. After eating crackers, *Ss* were told that there would be one other part of the experiment in which they would participate. Then they were requested to step into the hall to wait until *E* called them. The period was about three minutes. A stimulus sign was placed above the drinking fountain in the hall. The printed sign read, "Do not use this fountain." During the waiting period, the informed *S* either conformed to the sign, or violated it by drinking. For control conditions he was not present.

Induction of Thirst

Three degrees of thirst were induced. Intense thirst was created by feeding six crackers, three treated with Mexican hot sauce and belladonna. Three untreated crackers were used to produce a weak thirst. A puzzle problem was employed for evaluating frequency of drinking when thirst was not manipulated.

RESULTS

Violation with Degrees of Thirst

Inspection of row totals in Table 1 shows that violation increases with the conditions employed to increase thirst. Whereas none of the 20 *Ss* under the no thirst condition violated, 16 of 20,

TABLE 1

CRITICAL *Ss*' RESPONSES TO THE REACTIONS OF THE INFORMED PARTICIPANT AND TO VARIOUS DEGREES OF INDUCED THIRST

Degree of Thirst	Informed *S*				Total by Thirst	
	Drinks		Does Not Drink			
	Critical *S*		Critical *S*			
	Drinks	Does Not Drink	Drinks	Does Not Drink	Drinks	Does Not Drink
Strong	9	1	7	3	16	4
Weak	5	5	0	10	5	15
Non induced	0	10	0	10	0	20
Total by informed *S*	14	16	7	23		

or 80 per cent, violated under the strong thirst condition. The weak thirst condition, with 5 violations among 20 *Ss*, or 25 per cent, elicited a larger number of violations than the no thirst condition. The χ^2 of 25.65 for these differences (2 *df*) is significant far beyond the .01 level. Yates' correction for continuity was applied to χ^2 values.

Violation with Social Properties of Background

Column totals demonstrate that more violations occur when the *S* sees another person violating than when he sees another person conforming to the prohibition. Fourteen of 30 responses were violations when the confederate violated, whereas only seven of 30 were violations when the confederate conformed. The χ^2 for this difference is 3.62 and, with 1 *df*, approaches significance at the .05 level. Computed for the two thirst conditions, those under which violations would be expected to occur, the χ^2 is 4.80 and, with 1 *df*, approaches significance at the .01 level. The χ^2 of 7.20 (1 *df*) for the weak thirst condition, the condition in which *S* is neither intensely thirsty nor not thirsty, is significant beyond the .01 level. Under this condition, the action of the confederate has maximum influence on the *Ss*' behavior.

Violation as a Joint Function of Social Properties and Degree of Thirst

The χ^2 of 1.63 for the interaction between social background and strength motivation is not significant, indicating that there is no complex interrelationship between the two variables. The higher frequency of violations associated with increased thirst is independent of higher frequency associated with the behavior of the other person in the situation. Both factors, therefore, are important determinants of readiness to act against a prohibition.

SUMMARY

Ninety students who served as *Ss* understood that they were participating in an experiment in the field of taste, when the factor under investigation was behavior relative to a prohibition. The drive varied was thirst, the prohibited action was drinking water from a fountain, and the social background was perceived conformity or violation of the prohibition by another person. Behavior produced by a physiological drive of given strength is shown to be determined partly by the properties of the drive and partly by other properties of the situation which are social in character and which are not intrinsically related to the drive.

REFERENCES

1. FREED, A. M., CHANDLER, P. J., MOUTON, JANE S., & BLAKE, R. R. Stimulus and background factors in sign violation. *J. Pers.*, 1955, **23**, 499.
2. GROSSER, D., POLANSKY, N., & LIPPITT, R. A laboratory study of behavioral contagion. *Hum. Relat.*, 1951, **4**, 115.
3. LEFKOWITZ, M., Blake, R. R., & MOUTON, JANE S. Status factors in pedestrian violation of traffic signals. *J. abnorm. soc. Psychol.*, 1955, **51**, 704–705.

Received March 7, 1957.

THE ACHIEVEMENT MOTIVE AND TASK PERFORMANCE: A REPLICATION

ROBERT C. BIRNEY

Amherst College

THE findings reported here grew out of an effort to replicate previous research on the achievement motive (n Ach), in which a positive relationship had been reported between n Ach and simple speed tasks. Lowell (5) demonstrated a positive relationship between n Ach scores and a speed test of simple addition, while French (3) showed the same relationship using a digit-substitution task with a population of officer candidates. The intent of the study was to repeat as precisely as possible the original conditions of Lowell. The difficulties encountered in effecting this replication seem worthy of report.

METHOD

Four groups were used. Group 1 consisted of 24 male *Ss* who volunteered from the introductory class in psychology in which the investigator was an instructor. The investigator served as *E*. Group 2 consisted of 21 male *Ss* from the investigator's class, with another faculty member well known to the students as *E*.[1] For neither of these groups were Lowell's results confirmed. A re-examination of his original work brought out the fact that Lowell at the time of his study had been a graduate student experimenting with *Ss* from a sister institution. Similarly, French was apparently unknown to her *Ss* before the experiment. Accordingly, it seemed that a closer replication of Lowell's conditions demanded a student *E*. Group 3 consisted of 21 male *Ss* from one of the writer's classes, but the *E* was an honor student[2] in psychology. Group 4 consisted of 22 male *Ss* from the introductory course, again with the

[1] The writer is indebted to Professor Robert Grose for his assistance as *E* for this group.

[2] The writer is indebted to Gary Kirshbaum for his assistance as *E* for Groups 3 and 4.

TABLE 1

EFFECT OF EXPERIMENTER, INSTRUCTIONS, AND ACHIEVEMENT MOTIVE ON ADDITIONS AND DIGIT-SUBSTITUTION TASKS (Lowell's original findings included for comparison)

Group	Faculty *E*		Student *E*		Lowell (5)
	Group 1 (*N* = 24)	Group 2 (*N* = 21)	Group 3 (*N* = 21)	Group 4 (*N* = 22)	(*N* = 40)
Mean Additions					
High n Ach	146.5	133.3	157.8[a]	149.1	157.4[a]
N	14	11	11	9	19
Low n Ach	149.0	136.1	130.7	127.8	130.6
N	10	10	10	12	21
Mean Digit-Substitution					
High n Ach	260.6	218.9	281.0[a]	255.2[a]	
N	14	11	11	9	
Low n Ach	266.0	218.0	234.1	223.7	
N	10	10	10	12	
Total Group Means					
n Ach	5.54[b]	4.19[c]	+.57	−.27	
Additions	147.5	134.5	145.0	137.0	143.3
Digit-Substitution	262.7	218.5[d]	258.6	236.6	

[a] High n Ach performance reliably greater than Low's at .05 level.

[b] Reliably greater than Groups 3 and 4 at the .01 level.

[c] Reliably greater than Group 3 at the .05 level, Group 4 at the .01 level.

[d] Reliably less than Groups 1 and 3 at the .05 level.

same student acting as *E* but with a change in performance task instructions.

All groups were given the n Ach TAT in booklet form with standard instructions (6, p. 98).[3] After the TAT, the Additions task was administered followed by the Digit-Substitution task. The Additions task (summing three two-digit numbers) ran for ten minutes, one minute per page of twenty-five problems. The Digit-Substitution task ran for eight minutes, one minute per page of seventy letter-number combinations.

Groups 1, 2, and 3 received the following achievement orientation instructions after writing the stories and before attempting the tasks:

> The second portion of this session is devoted to performance on two tasks which are enlarged versions of two typical intelligence test items. Both are speed tests, and both represent abilities which have been consistently tested in IQ tests. My purpose is to compare your performance on them, and to take the first steps in research to see whether or not longer samples of these behaviors might be more diagnostic than the shorter versions now in use. Put your name on the booklets when you receive them. . . . Before we pass the first booklet out, however, let me say that unless you do your very best on these tasks I cannot make good use of the results, so please give it all you have.

Group 4 was told to read the printed instructions, which described the task and its timing, and then they were set to work ("neutral" instructions). "Neutral" instructions were used with Group 4 to explore the nature of the n Ach-performance task relationships under minimal cue conditions.

[3] The pictures used were: TAT 7BM, TAT 8 BM, the inventors (see p 100, 6), the printers, (used by Lowell), and the college class, #21, and office, #17, from Birney (1).

RESULTS

The results for all groups appear in Table 1. A direct comparison with Lowell's results is given. Such a comparison is impossible for French's data since her results are reported as mean improvement scores between control and experimental sessions. Two main findings appear. First, the mean n Ach score is significantly greater with the faculty *E*s than with the student *E*. Second, Group 3, which is the group closest to Lowell's original conditions, provides a precise replication of his results. In Group 4 high and low n Ach differ reliably on the Digit-Substitution task but not on the Additions. A comparison of total group means for the tasks show no group differences on the Additions task, but Group 2 drops reliably below the other groups on the Digit-Substitution task.

DISCUSSION

In regard to the effects of the *E*s upon the n Ach score, the faculty persons with whom the *S*s had associated in classroom (achievement) settings appear to have had the effect of arousing imagery in *S*s who were not motivated to perform well on the tasks. This interpretation is consistent with McClelland's theory of imagery determination, that the presence of achievement imagery in a story is due in part to the strength of the achievement cues in the story writing situation (6, p. 194). Had the higher n Ach scores for Groups 1 and 2 reflected greater motivation we would expect these two groups to outperform Groups 3 and 4 on the tasks, but they did not do so. In fact, the over-all level of performance on the tasks is comparable across the four groups, and data for the Additions task are consistent with the findings of Lowell.

The results for Group 3 constitutes the closest replication of Lowell's findings. Here a student *E* arouses far less achievement imagery, but the occurrence of such imagery predicts performance on the tasks successfully. Previous research relating n Ach to performance tasks of the type used here has not involved the use of a faculty *E* having close association with the *S*s. The possibility is thus suggested that an *E* effect has been present in many previous studies without sufficient appreciation of its role.

Two further aspects of the data warrant attention. First, Group 4 shows no drop in mean task performance despite "task" as opposed to "achievement arousal" instructions before the tasks. Similar findings have been reported by French (3) and Johnston (4). Apparently, for some populations the difference between aroused and task instruction conditions has little effect on

performance. Second, Group 2 shows a considerably lower mean performance on the Digit-Substitution task despite a substantial level of achievement imagery. This group was run at 11:00 A.M. on a sunny Saturday morning in May. Analysis of their Digit-Substitution task performance, which came toward the end of the hour-long test session, shows a mean performance of 8.6 letters below Group 1 on the first trial, and 24.4 letters below Group 1 on the last four trials, the sharp drop toward the end of the hour accounting for the lower over-all mean.

Two alternative explanations for the general findings were explored. The difference in n Ach scores from the faculty *Es* to the student *E* might be due to shift in scorer set, since these data were being processed as they were obtained. This possibility was checked by selecting seven *Ss* from each group, matched within one point on total n Ach score, scrambling the stories, and scoring them without knowledge of their source. Scorer reliability for the scores of the 28 *Ss* is +.86 (rho), indicating no large or systematic shifts in scorer set.

A second possibility is that the achievement imagery obtained by the faculty *Es* contains indices of anxiety or fear of failure which in turn might be related to the performance measures. Following Clark (2), an analysis was made of the affective signs (positive or negative) of the various categories of the n Ach scoring system. Clark's work suggested that persons characterized by fear of failure and achievement anxiety might tend to write stories dominated by negative categories. A careful analysis of the stories of Groups 1 and 2 show no category distributions that might imply greater anxiety.

The present results thus imply, first, that subsequent research in this area should give careful attention to the experimenter-subject relationship, and second, that this relationship may profitably be taken as an independent variable.

SUMMARY

Efforts to repeat the work of Lowell and French showing a positive relationship between achievement motivation and striving behavior are described. Two groups in which the *E* was a faculty member known to the students failed to reproduce the earlier results. Substitution of a student *E* resulted in good replication of the findings of Lowell and of French.

REFERENCES

1. BIRNEY, R. C. Studies on the role of picture cues in projective measurement of achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1955.
2. CLARK, R. A. A factor analytic integration of projective and clinical measures of need achievement. Unpublished master's thesis, Wesleyan Univer., 1948.
3. FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
4. JOHNSTON, R. A. The effects of achievement imagery on maze-learning performance. *J. Pers.*, 1955, **24**, 145–152.
5. LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1951, **33**, 31–40.
6. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Received March 14, 1957.

TWO COMPONENTS OF ASSUMED SIMILARITY[1]

ELIZABETH ALFERT

University of California

A MEASURE in social perception developed recently by Cronbach *et al.* (2) is the one of Assumed Similarity (AS) which is derived by comparing a person's self description with his prediction of the self description of others. Fiedler (3) has regarded AS as a measure of psychological distance, and has used it in connection with sociometric choices to investigate the structure and dynamics of social groups (e.g., 4). Rudin *et al.* (9) found AS scores to be consistent from one series of items to another and considered Assumed Similarity a measurable personality trait in which consistent individual differences exist. Cronbach (1) has called Assumed Similarity a general mental set almost independent of the psychological content of the items used, for which a global index would provide a satisfactory score.

This investigation is concerned with the processes involved in interpersonal perception. The possibility has been considered that the attribution of one's traits to others may not be a homogeneous mental set, but may involve two different predispositions: a predisposition to attribute traits in which ideal self and self coincide (ideal-congruent traits), and one to attribute traits in which ideal and self differ (ideal-discrepant traits).

[1] The author is greatly indebted to Dr. Egerton L. Ballachey who made this study possible by permitting the collection of data in his classes, and who has contributed much valuable advice and criticism.

If this could be established, Assumed Similarity would have to be considered a complex mental set, and an individual's global Assumed Similarity score would be a function of the relative number of ideal-congruent and ideal-discrepant items in the sample of items used.

In order to test whether Assumed Similarity scores are dependent on item content, the hypothesis is examined that *ideal-congruent traits and ideal-discrepant traits are attributed to others to a different degree*, by comparing individuals' mean Assumed Similarity scores for ideal-congruent traits (AS_c) with their mean Assumed Similarity scores for ideal-discrepant traits (AS_d). It may also be of interest to see whether individuals' AS_c scores can be predicted from their AS_d scores, and the relationship between AS_c and AS_d scores will, therefore, be investigated.

A person who attributes his ideal-congruent and ideal-discrepant traits to others in equal degree would have an AS_d/AS_c ratio of 1.00. A ratio smaller than 1.00 indicates that ideal-congruent traits are attributed to others to a greater extent than ideal-discrepant traits, while a ratio of more than 1.00 means that ideal-discrepant traits are attributed more than ideal-congruent traits. The present study also seeks to throw light on whether an individual's AS_d/AS_c ratio is indicative of a fairly consistent attitude and whether it is determined to any extent by his self satisfaction (the number of ideal-congruent traits). A further hypothesis, thus, holds that *there is a significant correlation between the amount of ideal-self congruency and the ratio of* AS_d/AS_c.

METHOD

Since this study is concerned with the influence of processes in the perceiver rather than with the accuracy of the perception of the behavior of the person perceived, unstructured stimuli, namely strangers were used. Gage (5) has shown that predicting the responses of strangers turned out to be less accurate than predicting the average. Judging strangers may be a more life-like situation and may thus reveal more of the motivational factors of the perceiver.

Two experimental groups of mixed sex were used: Group A, an adult extension class group of 27 people, and Group B, an undergraduate class group of 13 students. The students were unacquainted with each other at the time of the study. The stimulus persons to be described were members of the group, homogeneous as to race and social class, and they alternated in sex. Six stimulus persons were used for Group A, twelve persons for Group B.

The trait list used for Group B was constructed from Murray's set of personality variables (7), the 20 needs and five of the general attributes, choosing a trait name or attitude pertaining to a certain need and pairing it with an adjective of opposite meaning. A similar but less systematic list of 20 opposite traits had been used for Group A. Item pairs such as forceful—gentle, light-hearted—serious, spontaneous—self-controlled are examples of pairs used.

Each student in Group A and Group B was asked: (*a*) to check the trait list of opposite pairs, checking in each trait pair the one which would describe you better; (*b*) to check the list, according to your guess, as several specific individuals would check it for themselves; (*c*) to check the list descriptive of your self ideal, how you would ideally like to be.

In Group B, an additional task was added between steps 1 and 2. Each student was asked to check the list so as to describe, according to his guess, other specific individuals in the class.

Procedure for Scoring

AS has usually been scored by finding the difference between *S*'s self-description and *S*'s prediction of *O*'s self-description, but since a dichotomy is involved here, AS is expressed in terms of correspondences between *S*'s self-description and his prediction of *O*'s self-description.

1. AS_g, the global Assumed Similarity score was computed for each individual by summating correspondences between description of self and the predicted trait of the stimulus person for all items and for all persons, dividing by the number of stimulus persons and by the number of items.

2. An AS_c score for each individual was calculated by summating correspondences between self and the predicted trait of the stimulus person for all persons and for all ideal-congruent items, dividing by the number of stimulus persons and by the number of ideal-congruent items.

3. An AS_d score for each individual was computed by summating correspondences between self and the predicted trait of the stimulus person for all persons and for all ideal-discrepant traits, dividing by the number of stimulus persons and by the number of ideal-discrepant items. AS_g, AS_c and AS_d based on the description of *O* rather than on the prediction of *O*'s responses are similarly calculated.

AS_c can be considered as the AS score for the average ideal-congruent trait or the percentage of *O*s to whom the average ideal-congruent trait is attributed, while AS_d is the AS score for the average ideal-discrepant trait.

RESULTS

Ideal-self congruency in Group A ranged from 40 to 85 per cent, with a mean of 62 per cent and in Group B from 36 to 84 per cent, with a mean of 70 per cent. Agreement as to which member of the trait pair was the ideal varied from unanimous acceptance of one to about equally divided opinions, depending on the item pair.

Hypothesis 1. Ideal-congruent traits are attributed to others to a different extent than ideal-discrepant traits. Mean scores for AS_c and AS_d are given in Table 1. The difference between individuals' AS_c and AS_d was calculated and the significance of the mean difference tested. The difference between AS_c and AS_d is significant beyond the .001 level of confidence, in the predic-

tion of *O*'s self-description as well as in the description of *O*. (No significant difference at the .05 level was found for Group B between AS scores based on prediction and AS scores based on description.) An individual's global AS falls somewhere between his AS_e and AS_d score, approaching AS_e with greater ideal-self congruency. It thus seems evident that the psychological content of the item influences the extent to which similarity is assumed.

The possible influence of the general desirability of a trait (as determined by the opinion of the group) on the attribution of one's traits to others was also checked. Item pairs in Group A and B have been divided into two sets, 10 "extreme" item pairs for which the group was unanimous or almost unanimous as to which of the pair was the desirable trait, and 10 "neutral" item pairs for which less than two thirds of the group agreed as to the desirable trait. It was found that the average number of ideal-congruent items out of the ten was six in Group A for the extreme item pairs as well as for the neutral ones, and seven in Group B in extreme as well as in neutral items. Table 2 shows that AS rises with the general desirability of the trait. To some extent, though, the traits considered most unanimously as desirable may also be the ones among the ideal-congruent traits which are subjectively considered the most desirable ones. When the significance of the mean difference between AS_e and AS_d was tested in the neutral trait pairs, it was seen that the difference between the attribution of ideal-congruent and ideal-discrepant traits was still significant, $t = 3.15$ ($p < .01$) for Group A, and $t = 2.30$ ($p < .05$) for Group B.

A rank order coefficient of .08 between AS_e and AS_d was found for Group A and a rank order coefficient of .16 for Group B. The correlations are not significant at the .05 level. Though this finding does not prove that the variables are independent, it suggests that whatever association exists between them is weak.

Hypothesis 2. The AS_d/AS_e ratio was not significantly correlated with the amount of ideal-self congruency. A rank order coefficient of .10 between AS_d/AS_e and ideal-self congruency was found for Group A and a rank order coefficient of .18 for Group B. Plotted data show that people with high, medium, or low ideal-self congruency were equally likely to have a certain AS_d/AS_e ratio.

In order to examine whether the AS_d/AS_e ratio as well as AS_e and AS_d are indicative of fairly consistent attitudes within the individual, some of the reliabilities were checked. In Group A, split-half reliability was .64 for ideal-congruent items, .63 for ideal-discrepant items, across-person reliability was .57 for AS_e, .54 for AS_d and .58 for AS_d/AS_e. In Group B, slightly different item pairs were used, the number of item pairs increased from 20 to 25, and the number of stimulus persons increased from six to twelve. Across-item reliability was .86 for ideal-congruent items and .45 for ideal-discrepant items. Among the 25 items, the mean number of ideal-discrepant items was seven so that split-half reliability in the latter case could not have been expected to be very high. Across-person reliability for Group B was .97 for AS_e, .86 for AS_d and .93 for the AS_d/AS_e ratio. It seems that the AS_d/AS_e ratio is indicative of fairly consistent attitudes. Fig. 1 shows that the AS_d/AS_e distributions in groups A and B are of very similar appearance, with a mean of .72 for Group A, and .64 for Group B. Only three individuals out of 40 had ratios greater than 1.

TABLE 1

MEAN ASSUMED SIMILARITY SCORES FOR IDEAL-CONGRUENT AND IDEAL-DISCREPANT TRAITS

Group and Task	*N*	AS_e	AS_d	*t*	Level of Significance
Group A prediction	27	.64	.46	4.8	$p < .001$
Group B prediction	13	.74	.47	4.9	$p < .001$
Group B description	13	.71	.46	5.2	$p < .001$

TABLE 2

MEAN ASSUMED SIMILARITY SCORES FOR 10 EXTREME AND 10 NEUTRAL TRAIT PAIRS

Category of Trait	AS (Group A)	AS (Group B)
Ideal-congruent, unanimously considered desirable traits	.65	.82
Ideal-congruent traits of neutral items	.60	.65
Ideal-discrepant traits of neutral items	.49	.53
Ideal-discrepant, unanimously considered undesirable traits	.40	.33

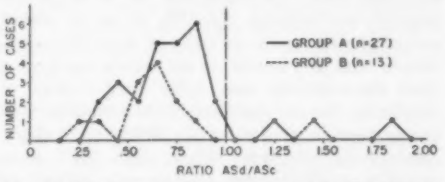

FIG. 1. FREQUENCY DISTRIBUTION OF INDIVIDUAL RATIOS OF AS_d/AS_e

DISCUSSION

There appears to be a general tendency for persons to perceive their ideal-congruent traits in

others to a greater extent than their ideal-discrepant traits. When individuals face unknown members of their group, they have a definite opinion as to how they compare to them and in general they feel themselves more alone in traits they consider as shortcomings.

Results based on the prediction of *O*'s responses alone might have been interpreted as *S*'s belief that *O* would be unlikely to attribute undesirable traits to himself, but since very similar results are derived regardless of whether AS is based on the description of *O* or on the prediction of *O*'s responses, it seems that *S* does not make much distinction between predicting *O*'s responses and perceiving *O*. Both probably are a function of perceiving *O* in some relation to the self image. The way others are perceived depends to a large extent on (*a*) the self image, the amount of ideal-self congruency or self-esteem, (*b*) the assumption that others are similar to oneself, (*c*) the linear relationship between desirability of the trait and the amount of assumption of similarity, and (*d*) the AS_d/AS_e ratio, the person's particular disposition to use different standards for attributing his ideal-congruent and his ideal-discrepant traits.

Since the stimulus is ambiguous, judgment of strangers could be considered as a projective method, and extreme AS_d/AS_e scores at either end might be taken as signs of maladjustment. People who are unable to see others as sharing their short-comings may suffer from feelings of inferiority. On the other hand, the rare cases who are extreme in predominantly seeing their short-comings in strangers rather than their good qualities, probably use this as ego defense or compensation mechanism. It seems likely that they felt the discrepancies between self and ideal so intensely that they had to compensate for them in order to maintain equilibrium. Klein and Schlesinger (6) have discussed the role of perception for maintaining equilibrium.

Zimmerman (10) showed that discrepancy between self and ideal may not indicate conflict. The present results for Hypothesis 2 indicated that ideal-self congruency does not influence the AS_d/AS_e ratio; it seems that traits discrepant from the ideal can be accepted by the ego, and only if such discrepancies become unacceptable to the ego do people attribute ideal-congruent traits to a lesser extent than ideal-discrepant traits in order to re-establish equilibrium. Rogers (8) sees adjustment related to the correspondence between self and ideal. Considering AS_d/AS_e scores in addition to ideal-self congruency might provide a better method for judging adjustment, since adding this measure would not only indicate a person's ideal-self discrepancies but also show how he reacts to them. It would be interesting to see whether extreme AS_d/AS_e scores are associated with certain maladjustments and whether therapy would revert such ratios towards the central tendency. Maladjustment has often been considered a disturbance in interpersonal relationship; how a person perceives others and how he perceives himself within the framework of his group might be a way of measuring it.

SUMMARY

1. Ideal-congruent traits are attributed to others to a greater extent than ideal-discrepant traits.

2. Assuming similarity for ideal-congruent traits and assuming similarity for ideal-discrepant traits are two separate attitudes and determine the global AS score.

3. Ideal-self congruency does not determine the relationship between attributing ideal-congruent and ideal-discrepant traits to others.

4. Central tendency and clinical implications of extreme cases are discussed.

REFERENCES

1. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assuming similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
2. CRONBACH, L. J., HARTMANN, W., & EHART, M. E. Investigation of the character and properties of Assumed Similarity measures. *Univer. of Illinois Bur. of Res. and Serv.*, 1953, Tech. Rep. No. 7, Contract 6 ori-07135.
3. FIEDLER, F. E. The psychological distance dimension in interpersonal relations. *J. Pers.*, 1953, **22**, 142–150.
4. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
5. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
6. KLEIN, G. S., & SCHLESINGER, H. J. Where is the perceiver in perceptual theory? *J. Pers.*, 1949, **18**, 32–47.
7. MURRAY, H. A. *Explorations in personality*. New York: Oxford Press, 1938.
8. ROGERS, C. R. The significance of the selfregarding attitudes and perceptions. In M. S. Reymert (Ed.), *Feelings and emotions*. New York. McGraw Hill, 1950.
9. RUDIN, S. A., LAZAR, I., EHART, M. E., & CRONBACH, L. J. Some empirical studies of the reliability of interpersonal perception scores. Tech. Rep. No. 4, Contract N6 ori-07135, Urbana, Ill.: Bur. Res. Serv., Univer. of Illinois, 1952.
10. ZIMMERMAN, H. Self acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, **18**, 447–449.

Received March 21, 1957.

NEUROTICISM, EXTRAVERSION, AND THE PERCEPTION OF THE VERTICAL[1]

RONALD TAFT
University of Western Australia

AND JAMES COVENTRY
New South Wales University of Technology

IN THEIR well-known studies of the relationship between personality factors and various tasks involving the perception of the upright (5), Witkin and his associates conclude that the relevant personality characteristics can "be reviewed under three headings: the nature of the individual's relation to his environment (including other people), the way in which he manages his impulses and strivings, and the kind of conception he has of himself" (p. 467). This present study deals with the relationship between (*a*) a simple test of ability to perceive the upright when, in the absence of a visual frame of reference, the angle of the body itself must be used as a reference, and (*b*) the personality traits Neuroticism and Extraversion-Introversion. These traits were selected by Eysenck (2) as representing the two main factors constituting the personality domain. Extraversion-Introversion appears to be similar to Witkin's first heading, and Neuroticism to his second, and perhaps his third. Let us now look at these in more detail.

Neuroticism. This trait, perhaps better referred to as emotional instability, implies inability to accept or to control one's impulses and strivings and lack of insight leading to anxiety, tension and self-deprecation. These characteristics suggest poor integration of the self, or "weak ego." Witkin's results indicate that persons who are high on these neurotic tendencies tend to be inaccurate in their perception of the upright. Witkin's findings are consistent with a common sense expectation that a neurotic person, deprived of the usual environmental cues upon which everyone depends in his normal life, would be less able to adapt himself to the use of somatic cues only, under the unusual circumstances of these experiments. Therefore, Hypothesis 1 is that *persons scoring high on a test of neuroticism perceive the upright less accurately than persons scoring low in neuroticism.*

Extraversion-Introversion. By extraversion we mean a personality orientation towards one's environment as opposed to social reserve, introspectiveness and self-sufficiency. This dimension is not quite the same as Witkin's "passivity-activity" relation to the environment, but we should expect the extravert to be more dependent on his environment, and therefore more "passive" in Witkin's sense than the introvert. Consequently, extraverts would be relatively unprepared to handle the environment by means of cues emanating from themselves rather than from the environment. Thus, Hypothesis 2 is that *persons scoring high on extraversion perceive the upright less accurately than those scoring high on introversion.*

METHOD

Personality Test

For convenience, a group-administered test, the Cattell 16 PF, was used to test neuroticism and extraversion-introversion: Factor C, "Emotional Stability or Ego-Strength versus General Neuroticism," was used to measure neuroticism and Factor H, "Adventurous Cyclothymic versus Inherent Withdrawn Schizothymic," to measure extraversion-introversion.[2] The test was scored in accordance with the norms published in the manual (1).

Spatial Orientation Apparatus

The equipment, consisting of a tilting chair and a luminous rod, was adapted from that described in (4). The chair was mounted on a length of tubing so that it could be turned in a vertical plane to the horizontal on either side, and the angle of tilt was measured by a protractor at the rear. The subject's position in the chair was kept constant by the use of shoulder pads, head rest, foot slots and a handle in the middle to keep the hands in a fixed position.

The rod was 39 inches long and situated five feet in front of the subject (*S*). From a handle operated behind the chair the rod could be turned in the vertical plane through any angle. It was illuminated by a series of small globes shining through a narrow slit along the center of the rod.

The testing was carried out in a dark room, the only source of light being the illumination from the rod.

Subjects and Experimental Set

The experimenter (*E*) administered the 16 PF to all Psychology I students in one of their regular laboratory

[1] This report is based on work done at the University of Western Australia. The authors thank Dr. A. J. Marshall for his assistance in the planning and administration of the perceptual testing.

[2] The traits included in the neurotic pole of Factor C are "lacking in frustration tolerance, changeable, showing general emotionality, evasive, neurotically fatigued, and worrying." On the extraverted pole of Factor H the traits are "gregarious sociability, adventurous, bold, having marked interest in the opposite sex, frivolous, strong artistic or sentimental interests, abundant emotional response."

TABLE 1

MEAN ERROR IN PERCEIVING THE VERTICAL WITH BODY TILTED AT VARIOUS ANGLES

Group	*N*	Tilted Right				Tilted Left			
		28°	42°	90°	Average	28°	42°	90°	Average
High C factor	26	3.7	6.0	10.0	6.6	3.6	6.8	11.0	7.0
Low C factor	21	3.5	6.0	10.2	6.7	2.8	5.3	9.4	5.8
t		—	—	—	—	1.01	1.15	1.01	1.30
High H factor	21	4.3	8.1	11.9	8.2	3.9	6.8	12.2	7.5
Low H factor	26	3.1	4.0	8.3	5.1	2.4	5.3	8.2	5.3
t		1.53	3.15**	2.14*	3.25**	2.30*	1.15	2.89**	2.52*

* Significant at 0.05 level.
** Significant at 0.01 level.

periods. They understood that they were taking this test both as part of their orientation to psychology and as part of their course obligation to act as *S*s for a number of student and staff research projects. No mention was made at this stage of a second part of the research. Unbeknown to the *S*s, only males were used and the tests were scored only on Factors C and H. The "middle" scorers (approximately the middle 50 per cent) on either factor were then discarded. The *E* then approached the remaining 47 men and sought an appointment for their participation in another experiment. There were no refusals. The *S*s were distributed as follows: high emotional stability (score 8–10 on C) 26, low emotional stability (1–5) 21; extraversion (7–10 on H) 21, introversion (1–3) 26. They were distributed in the four quadrants as follows:

High Stable—extravert	12
Low Stable—extravert	12
High Stable—introvert	14
Low Stable—introvert	9

Before taking *S* into the dark room, *E* explained that some work was being done on the perception of the upright, and that he was interested in how well people could judge the upright while they were in different body positions. All *S*s were told that nothing unexpected would happen to them, and were made to feel at ease before entering the dark room.

Procedure

In order to allay anxiety, *S*s were seated and "clamped" in the chair under full lighting conditions. The lights were then turned off until the completion of the testing. The *S*s were instructed as follows:

> I want you now to sit in this chair. You will be tilted at various angles in the vertical plane only. Place both hands on this handle and keep hold of it all the time. You need have no fear of falling as the chair has been specially weighted to counterbalance your body weight. See this rod in front of you? The rod lights up. When I switch it on I want you to tell me whether it is upright or not. By upright, I mean upright from the floor, not in relation to you. If it is upright say "It is upright." If it is not upright say "It is not upright."

Once these instructions were given, *E* seated *S* in the chair and adjusted the head rest and shoulder pads, making sure that *S* was as comfortable as he could be and that he had both hands placed firmly on the handle.

Judgments were then made by the *S* while tilted at the following angles:

(*a*) 0° Upright	(*e*) 28° Left
(*b*) 28° Right	(*f*) 42° Left
(*c*) 42° Right	(*g*) 90° Left
(*d*) 90° Right	(*h*) 0° Upright

In each of these body positions the rod was placed in an initial position 40° to the left of the upright and then adjusted by *E* until *S* was satisfied that it was vertical. Before *S* was tilted to a new angle these judgments were repeated three more times alternating the starting positions between 40° right and 40° left. The rod was moved from the starting positions in small increments, the light on the rod being switched off between each move.

On the average, the total time for each *S* was 30 minutes, with a maximum of 40 minutes.

Scoring

In each case the deviations from the vertical position were read to the nearest degree, disregarding sign. The deviations for the four judgments by each *S* were averaged in order to give a mean error score for every *S* at each of the eight positions.

RESULTS

In no case did the mean error score at 0° reach 1, so 0° is taken as the base for each *S*. The results for the other positions are summarized in Table 1. These results indicate no difference between the emotionally stable and unstable groups, and thus Hypothesis 1 is not confirmed. If anything, the unstable group are superior at the task than the stable. Hypothesis 2 is confirmed; the extraverts are consistently less accurate than the introverts.

When the *S*s falling into the four personality quadrants are compared, the nonneurotic introverts are the most accurate and the nonneurotic extraverts the least. The difference between these two extreme groups is significant at the 0.01 level for both the left and right tilts. The nonneurotic introverts are also significantly more accurate

than the neurotic extraverts (on the right tilt only).

DISCUSSION

The results for emotional stability appear to contradict those of Witkin and associates (5) and the common sense expectation that a person who is uncertain and in a state of conflict about himself in relation to his environment will be handicapped in using his own body as a framework for judging the environment. Why, then, did the less stable *Ss* perform as well as the more stable? It is impossible to answer this from the present data, but the discrepancy between our expectation and the findings could possibly be due to a mistaken opinion about the meaning of Factor C on the 16 PF test. In the manual for the test (1) the authors identify this factor with Eysenck's General Neuroticism, but it may not correspond with the characteristics described by Witkin and associates as "lack of awareness of inner life, fear of aggressive and sexual impulses, and poor control over these impulses . . . accompanied by . . . considerable anxiety, combined with difficulty in regulating it" (5, p. 468). It is possible that some persons "lacking awareness of their inner life" may appear more emotionally stable on a personal inventory such as the 16 PF than they really are.

Failure to obtain the expected results on neuroticism is unlikely to be explained in terms purely of the type of test used to perceive the upright. The task in our test appears to be psychologically equivalent to Witkin's tilting-room-tilting-chair Series 1-a and 1-b (5); i.e., it is a "field-as-a-whole task" in which the object of perception (the rod) cannot be directly related to the body. In these studies, poor performance on this type of test correlated significantly with a number of measures indicating neuroticism, including interview ratings of the *S*'s "anxiety versus self-assurance."

Apart from the possible misinterpretation of Factor C, we are at a loss to explain our results on neuroticism.

The confirmation of Hypothesis 2 requires little comment. Evidently, Factor H which we identified as extraversion is relevant to the ability to perceive the upright. Following Linton's analysis of the tilting-room-tilting-chair tests (3), we make the suggestion that Factor H is a measure of the "generalized tendency to accept or reject external influence."

SUMMARY

Forty-seven male *Ss* were tested on their ability to judge when a luminous rod was vertical, while they were tilted sideways in a chair within a dark room. It was hypothesized that persons scoring higher on tests of neuroticism and extraversion would be less accurate in perceiving the vertical than those scoring low on these traits. The traits were measured by the C and H scales respectively from the Cattell 16 PF questionnaire.

No difference was found between those scoring high or low on neuroticism, but those scoring high on extraversion were significantly less accurate in perceiving the vertical than were those scoring low.

REFERENCES

1. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. F. *The sixteen personality factor questionnaire.* Champaign, Ill. Institute for Personality and Ability Testing, 1950.
2. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
3. LINTON, HARRIET B. Dependence on external influence: correlates in perception, attitudes, and judgment. *J. abnorm. soc. Psychol.*, 1955, **5**, 502–507.
4. WITKIN, H. A., & ASCH, S. E. Studies in space orientation III. Perception of the upright in the absence of a visual field. *J. exp. Psychol.*, 1948, **38**, 603–614.
5. WITKIN, H. A. *et al. Personality through perception.* New York: Harpers, 1954.

Received August 9, 1956.

ON AUTHORITARIANISM AND ACQUIESCENCE: AN ADDED NOTE TO BASS AND MESSICK AND JACKSON

FRED N. KERLINGER

New York University

IN the May 1957 issue of this Journal, Messick and Jackson (2) criticized Bass's treatment and interpretation of the data of his study of authoritarianism and acquiescence (1). In their criticism Messick and Jackson stated that Bass's factor loadings have an incorrect sign. They also change the labels of the two factors Bass "extracted." This of course changes the interpretation of the table radically. I would like to add a bit more discord to this composition by transposing one of its themes to another key.

Neither Bass nor Messick and Jackson seemed to have noticed—at least they said nothing about it—what would happen if the two axes of the

alleged factors were rotated. Naturally, this is all in statistical fun since there is not much sense to the original factor analysis of one correlation coefficient. Two, three, or four rotations are equally possible and logical, depending on what you want the data to say. I want it to say just the opposite to Bass. So, if we rotate the axes 59 degrees clockwise using his data in (1), (the matrix on the left is Bass's original, unrotated; the one on the right is the rotated one) we get:

	Unrotated			Rotated	
	I Acquies.	II Author.		I′ Acquies.	II′ Author.
F	.77	.47	F	.01	.90
G	.53	−.47	G	.68	.21
				35%	65%

Now, instead of the statement by Bass that "... 59 per cent of the variance of the F scale is accounted for by Factor I, Acquiescence, while only the remaining 22 per cent of the reliable variance of the F scale concerns Authoritarianism as found in Factor II, "we can say that 35 per cent of the variance is associated with Acquiescence and 65 per cent with Authoritarianism!

Now suppose we play another game and rotate Messick and Jackson's correction of Bass's little table. Rotating these loadings 31 degrees counterclockwise (which corresponds to the rotation of Bass's original table given above), we get:

	Unrotated			Rotated	
	I Author.	II Acquies.		I′ Author.	II′ Acquies.
F	.77	.47	F	.90	.01
G	−.53	.47	G	−.21	.68
				65%	35%

With this "solution," we of course have the same interpretation as immediately above it, i.e., 65 per cent of the variance is associated with Authoritarianism and 35 per cent with Acquiescence, and approximately Messick and Jackson's interpretation of the unrotated matrix.

But notice what is common to both rotated matrices. The F scale is loaded only on one factor, presumably Authoritarianism. It has no loading at all on the other factor, presumably Acquiescence. The G scale, on the other hand, is loaded substantially on what is perhaps Acquiescence and negatively if only slightly loaded on Authoritarianism. This seems to land us almost back in the Adorno *et al.* lap! The analysis of one coefficient of correlation can get very complicated.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
2. MESSICK, S., & JACKSON, D. N. Authoritarianism or acquiescence in Bass's data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424–426.

Received September 11, 1957.

ARE THERE GAPS IN YOUR FILES OF APA JOURNALS?

SALE & BACK ISSUES

THEN HEAR THIS . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all *available* issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

American Psychologist
Journal of Abnormal & Social Psychology
Journal of Applied Psychology
Journal of Comparative & Physiological Psychology (1947–1950 only)
Journal of Consulting Psychology
Journal of Experimental Psychology
Psychological Abstracts
Psychological Bulletin
Psychological Index (a few complete volumes, some shopworn)
Psychological Monographs
Psychological Review

Not *all* issues in *all* volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery: 6 to 8 weeks • No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

Order from:

American Psychological Association
Department BB, 1333 Sixteenth Street, N.W.
Washington 6, D. C.

NOW AVAILABLE

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price $1.00

Discounts on quantity orders over 50 copies

Order from
AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office, 1333 Sixteenth Street, N.W., Washington 6, D. C.

An important new Wiley title . . .

EXPLORATIONS IN ROLE ANALYSIS

Studies of the School Superintendency Role

By NEAL GROSS, *Harvard University;* WARD S. MASON, *U. S. Office of Education; and* ALEXANDER W. McEACHERN, *The John Tracy Clinic, Los Angeles, California.* A compilation of the analyses of hundreds of interviews, this work brings social science theory to bear on important social science problems through the study of a single occupation. It also builds a "language" for role analysis that avoids the postulates which have blocked systematic role research in the past.

The authors deal with problem areas of central interest to readers concerned with social behavior. They present and test a theory of role conflict resolution based both on psychological and sociological variables. They offer one of the first systematic investigations of the impact of consensus on role definition, and give a penetrating treatment of conformity to expectations. Their book suggests many unexplored areas, emphasizing the crucial importance of consensus on role definition for cultural, social structural, and psychological analysis.

1958. 379 pages. $6.75

Send for an examination copy today

JOHN WILEY & SONS, Inc., 440 Fourth Avenue, New York 16, N.Y.